ANNALS OF MATHEMATICS
Vol. 38, No. 2, April, 1937

SINGULARITIES OF COMPOSITE FUNCTIONS IN SEVERAL VARIABLES

BY S. BOCHNER AND W. T. MARTIN[1]

(Received May 1, 1936)

In the present note we shall give a generalization to functions in several variables, of the following classical theorem of Hadamard.[2]

A. *Let*

$$a(z) = \sum_0^\infty a_n z^n, \qquad b(z) = \sum_0^\infty b_n z^n.$$

be any two power series converging in the neighborhood of the origin, and let $\pi(z)$ be the composite series

$$\pi(z) = \sum_0^\infty a_n b_n z^n.$$

If $\xi_\varphi e^{i\varphi}$, $\eta_\varphi e^{i\varphi}$, $\zeta_\varphi e^{i\varphi}$, $0 \leqq \varphi < 2\pi$, are the vertices of the stars of $a(z)$, $b(z)$, $\pi(z)$ respectively, then

$$\zeta_\varphi \geqq \underset{0 \leqq \vartheta < 2\pi}{\text{g.l.b.}} \{\xi_\vartheta \eta_{\varphi-\vartheta}\}.$$

The reader will find a proof of this classical theorem in an annex to this note.

The proof of the generalization will be simple enough but its formulation was by no means evident. In fact, the classical theorem itself will emanate in the following alternative form.

I. *Let*

$$a(x) = \sum_0^\infty a_n x^n, \qquad b(y) = \sum_0^\infty b_n y^n$$

be any two power series converging in the neighborhood of the origin, and let $\pi(x, y)$ be the composite series in two complex variables

$$\pi(x, y) = \sum_0^\infty a_n x^n \cdot b_n y^n.$$

If x^0, y^0 are any two numbers, then $\pi(x, y)$ is analytic along the segment

$$(\rho x^0, \rho y^0), \qquad 0 \leqq \rho^2 < \underset{0 \leqq \vartheta < 2\pi}{\text{g.l.b.}} \{\xi_\vartheta(x^0)\eta_{-\vartheta}(y^0)\},$$

[1] National Research Fellow.

[2] J. Hadamard, "Théorème sur les séries entières," *Acta Mathematica*, 22 (1898), 55–64; P. Dienes, *The Taylor Series*, Oxford (1931), esp. 346–348; S. Mandelbrojt, *Les singularités des fonctions analytiques representées par une Séries de Taylor*, Fascicule 54, Paris, (1932), 18–19.

where $\xi_\varphi(x^0)e^{i\varphi}$, $\eta_\varphi(y^0)e^{i\varphi}$ are vertices of the stars of the functions $a(x^0z)$, $b(y^0z)$, these functions being considered as functions of the complex variable z.

The transition from one to several variables will be based on the following key theorem of Hartogs.[3]

B. *Let $f(x_1, \cdots, x_\mu; z)$ be analytic in z in a fixed domain T containing the origin for every set $(x_1, \cdots, x_\mu)$ from the region*

$$I_\delta(x^0): |x_1 - x_1^0| + \cdots + |x_\mu - x_\mu^0| < \delta,$$

and analytic in $(x_1, \cdots, x_\mu; z)$ in the region $(I_\delta(x^0); |z| < r_0)$ for some $r_0 > 0$.

Then $f(x_1, \cdots, x_\mu; z)$ is analytic in $(x_1, \cdots, x_\mu; z)$ in the region $(I_\delta(x^0); T)$.

The corresponding generalization of Hurwitz's theorem[4] will involve notions that were introduced recently.[5]

Hadamard composite series. *Let $a(x_1, \cdots, x_\mu)$, $b(y_1, \cdots, y_\nu)$ be two power series, each in several complex variables. Grouping their terms in homogeneous polynomials, we put*

$$(1)\quad a(x_1, \cdots, x) = \sum_{n=0}^{\infty} A_n(x_1, \cdots, x_\mu), \qquad b(y_1, \cdots, y_\nu) = \sum_{n=0}^{\infty} B_n(y_1, \cdots, y_\nu).$$

Our composite series will be a function in all occurring variables,

$$(2)\qquad \pi(x; y) = \sum_{n=0}^{\infty} A_n(x_1, \cdots, x_\mu)\, B_n(y_1, \cdots, y_\nu).$$

We observe that this composition is invariant under any affine transformation in each group of variables. A composition of the coefficients themselves would fail in this and other respects.

Circular regions. We shall first make a very simple statement concerning the regions of uniform convergence of the three series (1), (2). These regions will be denoted by G_a, G_b, G_π respectively. Each of them is a circular region.[6] A region in the $(x_1, \cdots, x_\mu)$-space is circular if with any point $(x_1, \cdots, x_\mu)$ it contains all the points $(kx_1, \cdots, kx_\mu)$, $|k| \leqq 1$.

---

[3] F. Hartogs, "Zur Theorie der analytischen Funktionen mehrerer unabhängiger Veränderlichen, etc.," *Math. Annalen*, 62 (1906), 1–88, esp. 12–20.

[4] A. Hurwitz, "Sur un theorème de M. Hadamard," *Comptes Rendus*, Paris, 128 (1899), 350–353; S. Pincherle, "A priposita di un recente teorema del Sig. Hadamard," *Rendiconti de l'Acc. du Sc. de Bologn,a* 3 (1898–99), 67–74; S. Mandelbrojt, *loc. cit.*[2], 21–22.

[5] W. T. Martin, "Special regions of regularity of functions of several complex variables," *Proc. Nat. Acad. Sc.*, 22 (1936), 33–35.

[6] For a discussion of this and other fundamental notions see the tract by H. Behnke and P. Thullen, "Theorie der Funktionen mehrerer komplexen Veränderlichen," *Ergebnisse der Mathematik und ihrer Grenzgebiete*, Berlin (1934), 34, 41–47.

G_a is described by[7]

$$H_a(x_1, \cdots, x_\mu) < 1,$$

where

(3) $$H_a(x_1, \cdots, x_\mu) = \limsup_{x'=x} \left\{\limsup_{n=\infty} | A_n(x_1', \cdots, x_\mu') |^{1/n}\right\}.$$

The functions H_a, H_b, H_π satisfy the relation

$$H_\pi(x; y) = \limsup_{x'=x, y'=y} \left\{\limsup_{n=\infty} | A_n(x')B_n(y') |^{1/(2n)}\right\}$$

$$\leqq \{H_a(x_1, \cdots, x_\mu)H_b(y_1, \cdots, y_\nu)\}^{\frac{1}{2}},$$

hence:

II. *The composite function $\pi(x; y)$ is certainly analytic within*

$$H_a(x_1, \cdots, x_\mu)\, H_b(y_1, \cdots, y_\nu) < 1.$$

This is a circular region contained in the circular region G_π; indeed, the point $(x_1, \cdots, x_\mu; y_1, \cdots, y_\nu)$ belongs to G_π if numbers α, β with $| \alpha\beta | = 1$ exist such that $(\alpha x_1, \cdots, \alpha x_\mu)$ belongs to G_a and $(\beta y_1, \cdots, \beta y_\nu)$ belongs to G_b.

Stars. The star S_f of $f(x_1, \cdots, x_\mu)$ is defined as follows:[8] a point $(x_1^0, \cdots, x_\mu^0)$ belongs to S_f if $f(x)$ is regular along the segment $\rho x_1^0, \cdots, \rho x_\mu^0$, $0 \leqq \rho \leqq 1$. We shall need the following lemma which is based on Hartog's key theorem.

C. *Let $f(x_1, \cdots, x_\mu)$ be regular at the origin. Corresponding to any $x_1^0, \cdots, x_\mu^0$ let $\xi(x^0)$,*

$$0 < \xi(x^0) \leqq \infty,$$

be the vertex of S_f in the direction $x_1^0, \cdots, x_\mu^0$, that is, let $\xi(x^0)$ be the greatest number such that $f(x_1, \cdots, x_\mu)$ is regular along the segment $\rho x_1^0, \cdots, \rho x_\mu^0$, $0 \leq \rho < \xi(x^0)$. Then

(4) $$\xi(x^0) = \liminf_{x=x^0} \xi^*(x)$$

where $\xi^*(x)$ is the vertex on the positive real axis in the z-plane of the star of $f(x_1 z \cdots, x_\mu z)$, this function being considered as a function of z.

For the proof of this lemma, let ξ' denote any positive number less than $\liminf_{x=x^0} \xi^*(x)$ *for fixed* $x_1^0, \cdots, x_\mu^0$. There exists an $\epsilon = \epsilon(\xi'; x^0) > 0$ such that

$$\xi^*(x) \geqq \xi' \qquad \text{for } x_1, \cdots, x_\mu \text{ in } I_\epsilon(x^0).$$

---

[7] Compare W. T. Martin, "Special regions of regularity of functions of several complex variables." To appear in *Annals of Math.*

[8] See footnote 5 and B. Almer "Sur quelques problèmes de la théorie des fonctions analytiques de deux variables complexes," *Arkiv for Mathmatik, Astronomi och Fysik*, 17 (1922–23).

There exists a δ, depending only on ϵ and x^0 such that for any x from $I_\delta(x^0)$ and any real φ satisfying: $|\varphi| < \delta$ the new point $xe^{i\varphi}$ falls into $I_\epsilon'(x^0)$.

The vertices in the z-plane of the star of $f(x_1z, \cdots, x_\mu z)$ as a function of z are obviously $\xi^*(xe^{i\varphi})e^{i\varphi}$. Hence for x from $I_\delta(x_0)$ the function $f(x_1z \cdots, x_\mu z)$ is regular in the sector

$$|z| < \xi', \qquad |\arg z| < \delta.$$

On the other hand, for x from $I_\delta(x_0)$ and $|z|$ less some $r_0 > 0$ the function $f(x_1z, \cdots, x_\mu z)$ is regular in all variables. Defining T as the region consisting of the points of the sector

$$|z| < \xi', \qquad |\arg z| < \delta$$

and of the points of the circle $|z| < r_0$, we may now apply theorem B. We find that $f(x_1z, \cdots, x_\mu z)$ is analytic in $(I_\delta(x^0); T)$. Hence $f(x_1, \cdots, x_\mu)$ is analytic along the segment $\rho x_1^0, \cdots, \rho x_\mu^0$ for $\rho < \xi'$. Hence $\xi' \leqq \xi(x^0)$. But ξ' was any number smaller than the quantity on the right side of (4). Therefore

$$\xi(x^0) \geqq \liminf_{x=x^0} \xi^*(x).$$

But the greater sign cannot hold. Otherwise there would exist numbers ξ'' and ξ''' and a $I_\delta(x^0)$ such that

$$\xi(x^0) > \xi'' > \xi''' > \xi^*(x) \qquad \text{for } x \text{ from } I_\delta(x^0).$$

Since $f(x_1, \cdots, x_\mu)$ is regular along the segment

$$\overset{0}{x_1}\rho, \cdots, \overset{0}{x_\mu}\rho, \qquad \rho < \xi(x^0),$$

there exists and $I_{\epsilon'}(x^0)$ such that $f(x_1, \cdots, x_\mu)$ is regular along the segments

$$x_1\rho, \cdots, x_\mu\rho, \qquad \rho < \xi'', \qquad x \text{ from } I_{\epsilon'}(x^0).$$

Hence there exist an ϵ and a φ_0 such that $f(x_1z, \cdots, x_\mu z)$ is regular for x from $I_\epsilon(x^0)$ and z from the sector

$$|z| \leqq \xi''', \qquad |\arg z| \leqq \varphi_0.$$

Let $(x_1', \cdots, x_\mu')$ be a point common to $I_\delta(x^0)$ and $I_\epsilon(x^0)$. By the definition of $\xi^*(x')$ we have $\xi^*(x') \geqq \xi'''$ and this contradicts the assumption $\xi''' > \xi^*(x')$

This completes the proof of lemma C and the theorem proper will follow easily.

III. *Let*

$$a(x) = \sum_0^\infty A_n(x_1, \cdots, x_\mu), \qquad b(y) = \sum_0^\infty B_n(y_1, \cdots, y_\nu)$$

be any two power series, each in several complex variables, converging in the neighborhood of the origin, and let $\pi(x; y)$ be the composite series in all occurring variables,

$$\pi(x; y) = \sum_{n=0}^\infty A_n(x_1, \cdots, x_\mu)\, B_n(y_1, \cdots, y_\nu).$$

If $(x^0; y^0)$ is any composite point then $\pi(x; y)$ is analytic along the segment

$$(\rho x^0; \rho y^0), \qquad 0 \leqq \rho^2 < \underset{0 \leqq \vartheta < 2\pi}{\text{g.l.b.}} \{\xi(x^0 e^{i\vartheta})\, \eta(y^0 e^{-i\vartheta})\}$$

where

$$x^0 e^{i\varphi}\, \xi(x^0 e^{i\varphi}); \qquad y^0 e^{i\varphi}\, \eta(y^0 e^{i\varphi})$$

are vertices of the stars of the given functions $a(x_1, \cdots, x_\mu), b(y_1, \cdots, y_\nu)$.

The vertices of the stars of the functions

$$a(xz) = \sum_0^\infty A_n(x) z^n, \qquad b(yz) = \sum_0^\infty B_n(x) z^n,$$

as functions in z, may be written in the form

$$e^{i\varphi}\, \xi^*(xe^{i\varphi}), \qquad e^{i\varphi}\, \eta^*(ye^{i\varphi}).$$

By Hadamard's classical theorem (see A)

$$\sum_0^\infty A_n(x) B_n(y) w^n$$

is analytic along the segment

$$0 \leqq \Re(w) < \underset{0 \leqq \vartheta < 2\pi}{\text{g.l.b.}} \{\xi^*(xe^{i\vartheta})\eta^*(ye^{-i\vartheta})\}, \qquad \Im(w) = 0.$$

Hence

$$\pi(xz; yz) \equiv \sum_0^\infty A_n(x) B_n(y) z^{2n}$$

is analytic along the segment

$$z = \rho, \qquad 0 \leqq \rho^2 < \underset{0 \leqq \vartheta < 2\pi}{\text{g.l.b.}} \{\xi^*(xe^{i\vartheta})\eta^*(ye^{-i\vartheta})\}.$$

Introducing the quantities

$$\zeta(x^0; y^0), \qquad \zeta^*(x; y)$$

in analogy to $\xi(x^0)$, $\xi^*(x)$ and $\eta(y^0)$, $\eta^*(y)$ puts the last relation into the form

$$[\zeta^*(x; y)]^2 \geqq \underset{0 \leqq \vartheta \leqq 2\pi}{\text{g.l.b.}} \{\xi^*(xe^{i\vartheta})\eta^*(ye^{-i\vartheta})\}.$$

Since $\xi^*(xe^{i\vartheta})$ is for fixed x the vertex distance of a function in one variable, it is continuous from below in the variable ϑ. The same holds for $\eta^*(ye^{-i\vartheta})$, and consequently for $\xi^*(xe^{i\vartheta})\eta^*(ye^{-i\vartheta})$. But such a function attains a minimum on a closed set. Therefore we have

$$[\zeta^*(x; y)]^2 \geqq \xi^*(xe^{i\vartheta(x,y)})\eta^*(ye^{-i\vartheta(x,y)}).$$

There exists sequences $x^n \to x^0$; $y^n \to y^0$, $\vartheta_n = \vartheta(x^n, y^n) \to \vartheta_0$ such that

$$\liminf_{x;y=x^0;y^0} \xi^*(xe^{i\vartheta(x,y)})\eta^*(ye^{-i\vartheta(x,y)}) = \lim_{n=\infty} \xi^*(x^n e^{i\vartheta_n}) \cdot \lim_{n=\infty} \eta^*(y^n e^{-i\vartheta_n}).$$

In consequence of lemma C the last expression is

$$\geqq \xi(x^0 e^{i\vartheta_0})\cdot \eta(y^0 e^{-i\vartheta_0}).$$

But

$$\liminf_{x;\,y=x^0;\,y^0} \zeta^*(x;y) = \zeta(x^0;y^0),$$

hence

$$[\zeta(x^0, y^0)]^2 \geqq \xi(x^0 e^{i\vartheta_0})\cdot \eta(y^0 e^{-i\vartheta_0}).$$

Hence

$$[\zeta(x^0;y^0)]^2 \geqq \underset{0\leqq\vartheta<2\pi}{\text{g.l.b.}}\ \{\xi(x^0 e^{i\vartheta})\cdot \eta(y^0 e^{-i\vartheta})\}$$

and this is precisely the statement of our theorem.

The Hurwitz composition. With (1) we form

$$c(x;y) = \sum_{n=0}^{\infty} C_n(x;y) \tag{5}$$

where

$$C_n(x;y) = \sum_{k=0}^{n} \binom{n}{k} A_k(x) B_{n-k}(y).$$

If circular regions of convergence of our three series are characterized by $H_a(x) < 1$, $H_b(y) < 1$, $H_c(x, y) < 1$ then, compare (3),

$$H_c(x;y) \leqq H_a(x) + H_b(y). \tag{6}$$

A more refined relation, analogous to (6), is

$$h_c(x;y) \leqq h_a(x) + h_b(y) \tag{7}$$

where $h_a(x)$, $h_b(y)$, $h_c(x;y)$ describe the diagrams of the series $a(x)$, $b(y)$, $c(x;y)$. The relation follows from

$$h_a(x^0) = \limsup_{x=x^0}\left\{\limsup_{\rho=\infty} \frac{\log|A(x\rho)|}{\rho}\right\}$$

where

$$A(x) = \sum_{0}^{\infty} \frac{A_n(x)}{n!}$$

and from the analogous formulas for $h_b(y)$, $h_c(x;y)$, see footnote 5.

If $\alpha(x^0)$, $\beta(y^0)$, $\gamma(x^0;y^0)$ are the reciprocal values of the vertex distances of our three functions then

$$\gamma(x^0;y^0) \leqq \text{l.u.b. } \{\alpha(x^0), \beta(y^0), \alpha(x^0 e^{-i\varphi})e^{i\varphi} + \beta(y^0 e^{-i\psi})e^{i\psi}\}$$

taken over such real values of φ and ψ for which the expression in which they occur is positive.

A generalization. Let $k_1, \cdots, k_\mu; l_1, \cdots, l_\nu$ be any positive integers and suppose that the given expressions $A_n(x)$, $B_n(y)$ occurring in the Hadamard and Hurwitz composites are polynomials of weight n in the variables $x_1, \cdots, x_\mu$; $y_1, \cdots, y_\nu$, if these variables are given the weights $k_1, \cdots, k_\mu; l_1, \cdots, l_\nu$. Then all assertions remain valid, if the circular regions of uniform convergence are the corresponding Cartan circular regions pertaining to the indices $(k_1, \cdots, k_\mu)$, $(l_1 \cdots, l_\nu)$ and $(k_1, \cdots, k_\mu; l_1, \cdots, l_\nu)$ respectively and the stars and diagrams are similarly generalized, see footnote 5.

An illustration. Let $X_p(x)$, $p = 1, 2, \cdots$ denote a sequence of homogeneous polynomials in $x_1^{\kappa_1}, \cdots, x_\mu^{\kappa_\mu}$, let $\| X_p \|$ be the maximum of $| X_p(x) |$ for $| x_1 |^2 + \cdots + | x_\mu |^2 = 1$; and let α_p be a sequence of numbers. If the sequence $\| X_p \|$ is bounded and the series

$$\sum_{p=1}^{\infty} \frac{| \alpha_p |}{\| X_p \|}$$

is convergent, then the series

$$\sum_{p=1}^{\infty} \frac{\alpha_p}{1 - X_p(x)}$$

is uniformly absolutely convergent in the neighborhood of the origin, and defines a function which will be our $a(x)$. Its expansion $\sum_0^\infty A_n(x)$, where

$$A_n(x) = \sum_{p=1}^{\infty} \{X_p(x)\}^n$$

is its Cartan diagonal expansion for the indices

$$k_j = \frac{\kappa_1 \cdots \kappa_\mu}{\kappa_j} = \frac{\kappa}{\kappa_j}, \qquad j = 1, \cdots, \mu$$

and converges in the Cartan circular region

$$(8) \qquad | X_p(x) | < 1, \qquad p = 1, 2, \cdots,$$

see footnote 7. It does not converge in a larger Cartan region since it has a singularity on the curve in which the analytic manifold

$$x_1 = x_1^0 z^{k_1}, \cdots, x_\mu = x_\mu^0 z^{k_\mu}$$

intersects the boundary of (8); in fact, the function in z

$$a(x^0 z^k) = \sum_{p=1}^{\infty} \frac{\alpha_p}{1 - X_p(x^0) z^\kappa}$$

has a singularity on the corresponding circle in the z-plane.[9] Hence its Cartan region is

---

[9] See, for example, A. Pringsheim, "Über bemerkenswerte Singularitätenbildung bei gewissen Partialbruchreihen," *Sitzungsberichte der Bayerischen Akademie* (1927), pp. 145–164, where also various references are given.

$$\text{(9)} \qquad \underset{p}{\text{l.u.b.}} \mid X_p(x) \mid < 1.$$

Let

$$b(y) = \sum_{q=1}^{\infty} \frac{\beta_q}{1 - Y_q(y)} = \sum_{n=0}^{\infty} B_n(y_1, \cdots, y_\nu)$$

be a similar expression in the y-space with indices $\lambda_1, \cdots, \lambda_\nu; l_1, \cdots, l_\nu$; $\lambda = \lambda_1 \cdots \lambda_\nu$. Its Cartan region is

$$\mid Y_q(y) \mid < 1, \qquad q = 1, 2, \cdots,$$

or

$$\text{(10)} \qquad \underset{q}{\text{l.u.b.}} \mid Y_q(y) \mid < 1.$$

The Hurwitz composite $c(x; y)$ is the expression

$$\sum_{p,q=1}^{\infty} \frac{\alpha_p \beta_q}{1 - X_p(x) - Y_q(y)}.$$

If we add the assumption $\kappa = \lambda$ then it has properties analogous to those of $a(x), b(y)$. In particular, its Cartan region is

$$\underset{p,q}{\text{l.u.b.}} \mid X_p(x) + Y_q(y) \mid < 1.$$

In agreement with our previous assertion we have

$$\underset{p,q}{\text{l.u.b.}} \mid X_p(x) + Y_q(y) \mid \leqq \underset{p}{\text{l.u.p.}} \mid X_p(x) \mid + \underset{q}{\text{l.u.b.}} \mid Y_q(y) \mid,$$

the equality *not* holding as a rule. The Hadamard composite is

$$\pi(x; y) = \sum_{p,q=1}^{\infty} \frac{\alpha_p \beta_q}{1 - X_p(x) Y_q(y)},$$

the resulting Cartan region being

$$\text{(11)} \qquad \underset{p,q}{\text{l.u.b.}} \mid X_p(x) Y_q(y) \mid < 1.$$

Thus in this case we even have the equality

$$\underset{p,q}{\text{l.u.b.}} \mid X_p(x) Y_q(y) \mid = \underset{p}{\text{l.u.b.}} \mid X_p(x) \mid \cdot \underset{q}{\text{l.u.b.}} \mid Y_q(y) \mid;$$

in particular, the inequalities (9), (10) imply the inequality (11), in accordance with our general theorem.

The stars of our functions do not lend themselves to a similarly simple analytic representation. But the result relating to stars can be partially illustrated. In general, the star contains the Cartan region; in fact the radial distance $\rho(x)$ of the Cartan region and the vertex distance $\xi(x)$ of the star satisfy the relation

$$\rho(x) = \underset{\varphi}{\text{g.l.b.}}\ \xi(xe^{i\varphi}).$$

From the self explanatory argument

$$\begin{aligned}
\tau(x; y) &= \underset{\varphi}{\text{g.l.b.}}\ \zeta(xe^{i\varphi}; ye^{i\varphi}) \\
&= \underset{\varphi}{\text{g.l.b.}} \left\{ \underset{\vartheta}{\text{g.l.b.}}\ \xi(xe^{i\varphi}e^{i\vartheta})\eta(ye^{i\varphi}e^{-i\vartheta}) \right\} \\
&\geqq \underset{\vartheta}{\text{g.l.b.}} \left\{ \underset{\varphi}{\text{g.l.b.}}\ \xi(xe^{i\vartheta}e^{i\varphi})\ \underset{\varphi}{\text{g.l.b.}}\ \eta(ye^{-i\vartheta}e^{i\varphi}) \right\} \\
&\geqq \underset{\vartheta}{\text{g.l.b.}}\ \{\rho(xe^{i\vartheta})\sigma(ye^{-i\vartheta})\} \\
&\geqq \rho(x)\sigma(y)
\end{aligned}$$

it follows that a point on the boundary of the Cartan region of $\pi(x; y)$ cannot be singular (that is a star vertex) unless it is a composite of such points of the functions $a(x)$, $b(y)$. And this statement can be verified for the Cartan regions (9), (10), (11).

To return to the Hurwitz composite if the constants $\kappa_1, \cdots, \kappa_\mu; \lambda_1, \cdots, \lambda_\nu$ are all 1, then the indicator functions $h_a(x)$, $h_b(y)$, $h_c(x; y)$ have the values

$$\begin{aligned}
h_a(x) &= \underset{p}{\text{l.u.b.}}\ \{\text{real part } X_p(x)\} \\
h_b(y) &= \underset{q}{\text{l.u.b.}}\ \{\text{real part } Y_q(y)\} \\
h_c(x; y) &= \underset{p,q}{\text{l.u.b.}}\ \{\text{real part } [X_p(x) + Y_q(y)]\},
\end{aligned}$$

and hence precisely

$$h_c(x; y) = h_a(x) + h_b(y).$$

Annex. Proof of theorem A. Let $e^{i\varphi}\rho(\varphi)$ denote the star vertices of a function $f(z)$ which is regular around the origin. Since $f(z)$ is regular along the segment $z = e^{i\varphi}\rho$, $0 \leqq \rho < \rho(\varphi)$ it follows that $\rho(\varphi)$, as a function of φ, is continuous from below.

Since both given functions ξ_ϑ, $\bar{\eta}_{\varphi-\vartheta}$ are continuous from below, their product, as a function in ϑ, is also continuous from below. Such a function attains its minimum on every closed set. Let $\alpha_n(\varphi)$, $\beta_n(\varphi)$ be sequence of continuous (periodic) functions converging increasingly towards ξ_φ, $\bar{\eta}_\varphi$ as $n \to \infty$. Then, for every φ the sequence $\alpha_n(\varphi)\beta_n(\varphi - \vartheta)$ converges increasingly towards $\xi_\vartheta\bar{\eta}_{\varphi-\vartheta}$. Since the functions of the sequence are continuous and converge increasingly, and their limit function attains its minimum, we have

$$\lim_{n=\infty} \underset{0 \leqq \vartheta < 2\pi}{\text{g.l.b.}}\ \{\alpha_n(\vartheta)\beta_n(\varphi - \vartheta)\} \geqq \underset{0 \leqq \vartheta < 2\pi}{\text{g.l.b.}}\ \{\xi_\vartheta\bar{\eta}_{\varphi-\vartheta}\}.$$

The appromating functions $\alpha_n(\varphi)$, $\beta_n(\varphi)$ can be so chosen that the stars $e^{i\varphi}\alpha_n(\varphi)$, $e^{i\varphi}\beta_n(\varphi)$ are bounded by rectifiable curves. Hence it is sufficient to prove the following statement.

If $a(z)$ is analytic in the closed star $e^{i\varphi}\alpha(\varphi)$ and $b(z)$ in the closed star $e^{i\varphi}\beta(\varphi)$ and the star boundaries are rectifiable then $\pi(z)$ is analytic in the interior of the composite star

$$z = e^{i\varphi}\rho \qquad 0 \leqq \rho < \min_{0\leqq\vartheta<2\pi} \{\alpha(\vartheta)\beta(\varphi - \vartheta)\}. \tag{12}$$

If t is a point of the contour $e^{-i\vartheta}/\beta(\vartheta)$,—we shall denote this contour by C,—and z is a point in (12) then the contour zt is interior to the star $e^{i\psi}\alpha(\psi)$. In fact, zt has the form $e^{i(\varphi-\vartheta)}\rho/\beta(\vartheta)$ and $\rho/\beta(\vartheta)$ is less than $\alpha(\varphi - \vartheta)$. Hence we can form the integral

$$\frac{1}{2\pi i}\int a(tz)b\left(\frac{1}{t}\right)\frac{dt}{t} \tag{13}$$

for all z in (12). Furthermore if z runs over a closed set S interior to (12) and t runs over C, then the totality of points tz is closed, and thus a closed subset of the interior of the star in which $a(z)$ is being considered. Therefore for all t on C and z from S, $a(tz)$ is uniformly continuous in (t, z) and analytic in z. Hence the function (13) is analytic in (12). Let α denote the minimum of $\alpha(\varphi)$ and β the minimum of $\beta(\varphi)$. Then $a(tz)$ is analytic for $|t| \leqq \alpha/|z|$ and $b(1/t)$ for $|t| \geqq 1/\beta$. Let $|z|$ be so small that C lies in the annulus $(1/\beta, \alpha/|z|)$. Then, in the integral (13), C can be replaced by any concentric circle in the annulus. On this circle we can substitute $\sum_0^\infty a_n t^n z^n$ for $a(tz)$ and $\sum_0^\infty b_n t^{-n}$ for $b(1/t)$, and thus we identify the integral (13) with the function $\pi(z)$.

PRINCETON, N. J.

ANNALS OF MATHEMATICS
Vol. 38, No. 2, April, 1937

THE POLYNOMIAL $F_n(x)$ AND ITS RELATION TO OTHER FUNCTIONS

BY H. BATEMAN

(Received December 14, 1936)

1. The function $U_n(z)$, which was used in a recent paper,[1] may be defined by means of the equation

$$F_n\left(\frac{d}{dx} - 1\right) \exp(-te^{-2x}) = U_n(te^{-2x}) \tag{1}$$

which is easily generalized. We may, for instance, write

$$F_m\left(\frac{d}{dx} - 1\right) F_n\left(\frac{d}{dx} - 1\right) \exp(-te^{-2x}) = U_{mn}(te^{-2x}) \tag{2}$$

and obtain a power series

$$U_{mn}(z) = \sum_{s=0}^{\infty} \frac{(-z)^s}{s!} F_m(-2s - 1) F_n(-2s - 1). \tag{3}$$

There is a similar generalization with any number of factors of type $F_m\left(\frac{d}{dx} - 1\right)$.

The function $U_{mn}(z)$ is related to the function $P_{m,n}(u)$ used in a former paper,[2] in fact

$$\frac{1}{1+x} P_{m,n}\left(\frac{x-1}{x+1}\right) = \int_0^\infty e^{-xt} U_{mn}(t)\, dt \qquad R(x) > 0 \tag{4}$$

and we may write for $|x| > 1$,

$$\frac{1}{1+x} P_{m,n}\left(\frac{x-1}{x+1}\right) = \sum_{s=0}^{\infty} (-)^s x^{-s-1} F_m(-2s-1) F_n(-2s-1), \tag{5}$$

while for $|x| < 1$

$$\frac{1}{1+x} P_{m,n}\left(\frac{x-1}{x+1}\right) = \sum_{s=0}^{\infty} (-x)^s F_m(2s+1) F_n(2s+1). \tag{6}$$

Since $F_n(-2s - 1) \geqq 1$ it follows from (5) that when $z > 1$, $P_{m,n}(z) > 1$.

2. By means of equation (1) or the relation $U_n(t) = \exp(-t) Z_n(t)$, where $Z_n(t)$ is the polynomial $F(-n, n+1; 1, 1; t)$ it is easily shown that

---

[1] H. Bateman, Duke Mathematical Journal, vol. 2 (1936) 569–577.
[2] H. Bateman, Annals of Mathematics, vol. 35 (1934) 767–775.

(7) $$\int_0^\infty \tau^{\nu-1} U_n(\tau)\, d\tau = \Gamma(\nu) F_n(2\nu - 1).$$

The associated equation

(8) $$\int_0^\infty \tau^{\nu-1} U_{mn}(\tau)\, d\tau = \Gamma(\nu) F_m(2\nu - 1) F_n(2\nu - 1)$$

may be proved by means of the equation

(9) $$U_{mn}(\tau) = e^{-\tau} \sum_{s=0}^{m+n} a_{m,n,s}\, \tau^s$$

in which the coefficients $a_{m,n,s}$ are those of the expansion

(10) $$F_m(2\nu - 1) F_n(2\nu - 1) = \sum_{s=0}^{m+n} a_{m,n,s}\, \nu(\nu + 1) \cdots (\nu + s - 1)$$

which gives (9) when it is used to evaluate the left hand side of (2). The relation (9) shows clearly the nature of the entire function $U_{mn}(z)$ and a more general function such as $U_{lmn}(z)$ is readily seen to be also equal to a polynomial multiplied by e^{-z}.

From (6) and (7) we infer that

(11) $$\int_0^\infty U_m(x\tau) U_n(\tau)\, d\tau = \frac{(-)^m}{1 + x} P_{m,n}\left(\frac{x - 1}{x + 1}\right).$$

A proof of this equation may be obtained from

(12) $$\int_0^\infty e^{-x\tau} U_n(\tau)\, d\tau = \frac{1}{1 + x} P_n\left(\frac{x - 1}{x + 1}\right)$$

by writing e^{-2x} in place of x and using (1). We find in a similar way that

(13) $$\int_0^\infty U_l(x\tau) U_{mn}(\tau)\, d\tau = \frac{(-)^l}{1 + x} P_{l,m,n}\left(\frac{x - 1}{x + 1}\right)$$

where

(14) $$\operatorname{sech} \omega \cdot P_{l,m,n} (\tanh \omega) = F_l\left(\frac{d}{d\omega}\right) F_m\left(\frac{d}{d\omega}\right) F_n\left(\frac{d}{d\omega}\right) \operatorname{sech} \omega.$$

Also, if

(15) $$F_l\left(\frac{d}{dx} - 1\right) F_m\left(\frac{d}{dx} - 1\right) F_n\left(\frac{d}{dx} - 1\right) e^{-te^{-2x}} = U_{lmn}(te^{-2x})$$

we have

(16) $$\int_0^\infty e^{-x\tau} U_{lmn}(\tau)\, d\tau = \frac{1}{1 + x} P_{l,m,n}\left(\frac{x - 1}{x + 1}\right)$$

and so on.

3. The function $V_n(x)$, orthogonal to $U_n(x)$, may be defined by means of the equation

$$F_n\left(\frac{d}{dx}-1\right)\int_1^\infty \exp\left[-t(\tau-1)e^{-2x}\right]\frac{d\tau}{\tau} = V_n(te^{-2x}) \tag{17}$$

for the left hand side is equal to

$$\int_1^\infty U_n[t(\tau-1)e^{-2x}]\frac{d\tau}{\tau} = \int_0^\infty \frac{U_n(s)\,ds}{s+te^{-2x}} \tag{18}$$

and so the present definition agrees with the former definition.

It is readily seen that

$$\int_0^\infty e^{-x\tau}V_n(\tau)\,d\tau = \frac{2}{x-1}\,Q_n\left(\frac{x+1}{x-1}\right), \qquad R(x) > 0 \tag{19}$$

and so, on replacing x by e^{-2x} and using (1), we find that

$$\int_0^\infty U_m(x\tau)V_n(\tau)\,d\tau = (-)^m\,\frac{2}{x-1}\,Q_{m,n}\left(\frac{x+1}{x-1}\right), \qquad R(x) > 0. \tag{20}$$

The orthogonal relation is now obtained by making $x \to 1$ and using a result obtained in [2].

Defining the function $V_{mn}(x)$ by the equation

$$V_{mn}(x) = \int_0^\infty \frac{U_{mn}(t)\,dt}{x+t} \tag{21}$$

or the equivalent equation

$$F_m\left(\frac{d}{dx}-1\right)F_n\left(\frac{d}{dx}-1\right)\int^\infty \exp\left[-t(\tau-1)e^{-2x}\right]\frac{d\tau}{\tau} = V_{mn}(te^{-2x}) \tag{22}$$

we have also

$$\int_0^\infty e^{-x\tau}V_{mn}(\tau)\,d\tau = \frac{2}{x-1}\,Q_{m,n}\left(\frac{x+1}{x-1}\right). \tag{23}$$

The last equation gives a rule for the determination of the coefficient a_n in an expansion

$$f(t) = \sum_{n=0}^\infty a_n V_n(t), \qquad t > 0. \tag{24}$$

After some reductions the rule is found to be

$$a_n = (2n+1)\int_0^\infty e^{-t}\,dt\left[F_n\left(1+2t\frac{d}{dt}\right)f(t)\right] \tag{25}$$

and is an alternative to the rule

$$a_n = (2n+1)\int_0^\infty U_n(t)f(t)\,dt. \tag{26}$$

In the particular case when $f(t) = t^{\nu-1}$ both rules give [for $R(\nu) > 0$]

$$a_n = (2n+1)\Gamma(\nu)F_n(2\nu - 1). \tag{27}$$

4. Defining the function $U^{m,n}(z)$ by the relation

$$F_{m,n}\left(\frac{d}{dx} - 1\right)e^{-te^{-2x}} = U^{m,n}(te^{-2x}) \tag{28}$$

in which $F_{m,n}(x)$ is the function introduced in [2] we have the power series

$$U^{m,n}(z) = \sum_{s=0}^{\infty} \frac{(-z)^s}{s!} F_{m,n}(-2s-1) \tag{29}$$

and the definite integral

$$\int_0^\infty e^{-x\tau} U^{m,n}(\tau)\,d\tau = \frac{1}{1+x} P_m\left(\frac{x-1}{x+1}\right) P_n\left(\frac{x-1}{x+1}\right). \tag{30}$$

This indicates that for $|x| > 1$

$$\frac{1}{1+x} P_m\left(\frac{x-1}{x+1}\right) P_n\left(\frac{x-1}{x+1}\right) = \sum_{s=0}^{\infty} (-)^s x^{-s-1} F_{m,n}(-2s-1) \tag{31}$$

and there is a corresponding expansion for $|x| < 1$. We have also

$$\int_0^\infty \tau^{\nu-1} U^{m,n}(\tau)\,d\tau = \Gamma(\nu)F_{m,n}(2\nu-1). \tag{32}$$

We may further define a function $V^{m,n}(z)$ by the equation

$$F_{m,n}\left(\frac{d}{dx} - 1\right)\int_1^\infty e^{-t(\tau-1)e^{-2x}} \frac{d\tau}{\tau} = V^{(m,n)}(te^{-2x}), \tag{33}$$

it has the property

$$\int_0^\infty e^{-x\tau} V^{(m,n)}(\tau)\,d\tau = \frac{1}{x-1} Q_m\left(\frac{x+1}{x-1}\right) P_n\left(\frac{x+1}{x-1}\right) \qquad m \geqq n \tag{34}$$

A generating function for $U^{m,n}(t)$ is readily found from (30), we have for $|y| < 1$, $|z| < 1$

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} y^m z^n U^{m,n}(t) = \frac{1}{(1-y)(1-z)} e^{-at} I_0(bt), \tag{35}$$

$$\text{where } a + b = \left(\frac{1+y}{1-y}\right)^2, \qquad (a-b)\left(\frac{1+z}{1-z}\right)^2.$$

If $U^{m,n}(t) = e^{-t}Z^{m,n}(t)$, the exponential factor in the generating function of $Z^{m,n}(t)$ is

$$\exp\left[-\frac{2yt}{(1-y)^2} - \frac{2zt}{(1-z)^2}\right]$$

and it is readily seen that $Z^{m,n}(t)$ is a polynomial of degree $m + n$ in t.

Putting $y = z$ in the generating function we have the expansion

$$\frac{1}{(1-z)^2} e^{-t\left(\frac{1+z}{1-z}\right)^2} = \sum_{n=0}^{\infty} z^n T_n^*(t) \tag{36}$$

where

$$\begin{aligned} T_n^*(t) &= \sum_{s=0}^{n} U^{n-s,s}(t) = e^{-t} \sum_{m=0}^{n} \frac{(-4t)^m (n+m+1)!}{m!\,(n-m)!\,(2m+1)!} \\ &= \frac{n+1}{2\pi i} e^{-t} \int_C \left(1 - \frac{4t}{z}\right)^n {}_1F_1\left(n+2; \frac{3}{2}; \frac{z}{4}\right) \frac{dz}{z}, \end{aligned} \tag{37}$$

and C is a simple contour enclosing the origin. It is readily seen that

$$\int_0^\infty e^{-xt} T_n^*(t)\, dt = \frac{1}{1+x} \frac{\sin (n+1)\theta}{\sin \theta}, \qquad x > 0, \tag{38}$$

where

$$\cos \theta = \frac{x-1}{x+1}.$$

The asterisk in $T_n^*(t)$ has been introduced to distinguish it from the Tchebycheff Polynomial $T_n(t)$.

On account of the orthogonal property of the Legendre polynomial it follows from (30) that

$$\begin{aligned} \int_0^\infty V_0(\tau)\, U^{m,n}(\tau)\, d\tau &= 0 \quad \text{if} \quad m \neq n \\ &= \frac{1}{2n+1} \quad \text{if} \quad m = n. \end{aligned} \tag{39}$$

A more general relation

$$\int_0^\infty V_s(\tau) U^{m,n}(\tau)\, d\tau = \frac{1}{2r+1} \frac{A_{r-n} A_{r-m} A_{r-s}}{A_r} \tag{40}$$

$$2r = m + n + s, \qquad s!\, A_s = 1\cdot 3\cdot 5 \cdots (2s-1)$$

may be deduced from the well known expansion (for $m \leqq n$)

$$\begin{aligned} P_m(z) P_n(z) &= \sum_{s=0}^{m} C_s P_{m+n-2s}(z), \\ (2m + 2n - 2s + 1) C_s A_{m+n-s} &= A_{m-s} A_{n-s} A_s \end{aligned} \tag{41}$$

and the related expansion

$$U^{m,n}(t) = \sum_{s=0}^{m} C_s U_{m+n-2s}(t). \tag{42}$$

The relations (37) and (39) give

$$(43)\qquad \int_0^\infty V_0(t)T_n^*(t)\,dt = (1/n+1) \quad \text{for} \quad n \text{ even}$$

$$= 0 \quad \text{for} \quad n \text{ odd}.$$

This is a particular case of a more general relation

$$(44)\qquad \int_0^\infty V_m(t)T_n^*(t)\,dt = \tfrac{1}{2}\int_0^\pi P_m(\cos\theta)\sin(n+1)\theta\,d\theta$$

which may be deduced from (19).

(39) and (35) give the relation

$$(45)\qquad \int_0^\infty e^{-at}I_0(bt)V_0(t)\,dt = 2[(r^2-1)(s^2-1)]^{-\frac{1}{2}}\log\frac{(r+1)(s+1)+\sqrt{(r^2-1)(s^2-1)}}{(r+1)(s+1)-\sqrt{(r^2-1)(s^2-1)}}$$

$$r^2 = a+b, \qquad s^2 = a-b, \qquad a \geqq b \geqq 0.$$

A series may be obtained for the corresponding integral with $V_n(t)$ in place of $V_0(t)$ but it is rather complicated.

It may be shown by means of (30) that

$$(46)\qquad U^{m,n}(t) = e^{-t}\left[1 - c_1\frac{t}{(1!)^3} + c_2\frac{t^2}{(2!)^3} - \cdots\right]$$

where

$$(47)\qquad \begin{aligned} c_1 &= m(m+1)+n(n+1).\\ c_2 &= (m-1)m(m+1)(m+2)+4m(m+1)n(n+1)+(n-1)\\ &\quad n(n+1)(n+2).\\ c_3 &= (m-2)(m-1)\,m(m+1)(m+2)(m+3)+9(m-1)\,m(m+1)\\ &\quad (m+2)\,n(n+1)+9(m(m+1)(n^2-n)(n+1)(n+2)+(n-2)\\ &\quad (n-1)\,n(n+1)(n+2)(n+3). \end{aligned}$$

The related expansion for $F_{m,n}(-1\;-2x)$ is

$$(48)\qquad F_{m,n}(-1-2x) = 1 + c_1\frac{x}{(1!)^3} + c_2\frac{x(x-1)}{(2!)^3} + c_3\frac{x(x-1)(x-2)}{(3!)^3} + \cdots.$$

It should be mentioned that the functions $Z_n(t)$, $T_n^*(t)$ satisfy the recurrence relations

$$(49)\qquad (4n+2)\int_0^t Z_n(s)\,ds = (2n+1)Z_n(t) - (n+1)Z_{n+1}(t) - nZ_{n-1}(t)$$

$$\left(\frac{d}{dt}+1\right)[T^*_{n+1}(t)+T^*_{n-1}(t)] = 2\left(\frac{d}{dt}-1\right)T^*_n(t) \tag{50}$$

$$\begin{aligned} &2t\frac{d}{dt}[T^*_n(t)-T^*_{n-1}(t)] \\ &\quad = (n+1)T^*_{n+1}(t) + (2t-2n-2)T^*_n(t) + (n+1+2t)T^*_{n-1}(t). \end{aligned} \tag{51}$$

When the Neumann series

$$M(x^2t) = x^{-1}\sum_{n=0}^{\infty}(4n+2)Z_n(t)J_{2n+1}(x) \tag{52}$$

is arranged according to powers of x it is found that

$$M(x^2t) = \sum_{m=0}^{\infty}\frac{(-x^2t)^m}{2^{2m}(m!)^3} = f(-\tfrac{1}{4}xt, 0, 0) \tag{53}$$

in the notation of Wrinch.[3] This is an entire function of x^2t and so the Neumann series may be expected to converge for all finite values of x. The associated power series is

$$\sum_{n=0}^{\infty}\frac{x^{2n}}{2^{2n}(2n)!}Z_n(t)$$

and this must represent an entire function of x. If we retain only the highest power of t in $Z_n(t)$ so as to obtain a function of x^2t we get simply the power series for $M(x^2t)$ as associated series.

The series (46) suggests the formula

$$Z_n(t) = G\int_0^{\infty} M(x^2t)J_{2n+1}(x)\,dx, \tag{54}$$

the general value of the integral being obtained by introducing an exponential factor e^{-ux} under the integral sign and passing to the limit $u \to 0$. If it is calculated by expanding the function $M(x^2t)$ in a power series and associating the value $2^{2m}\Gamma(n+m+1)/\Gamma(n-m+1)$ with the divergent integral

$$\int_0^{\infty} x^{2m}J_{2n+1}(x)\,dx$$

the power series for $Z_n(t)$ is obtained.

---

[3] D. Wrinch, Quart. J. 50 (1927) 204–224. An asymptotic expansion for $M(x)$ may be derived from the general formulae of Fox, Proc. London Math. Soc. (2) 27 (1928) 389–400. The function M is related also to functions used by L. Pochhammer, Math. Ann. 41 (1893) 197–218 and A. R. Forsyth, Quart. J. 50 (1924) 97–148. Forsyth gives the asymptotic value $f(-z, 0, 0) \sim A \exp(\tfrac{3}{2}z^{\frac{1}{3}}) \cos(\tfrac{3}{2}\sqrt{3}\cdot z^{\frac{1}{3}} - \tfrac{1}{3}\pi)$ where A is a constant which he leaves undetermined.

It should be noticed that the integral

$$\int_0^\infty J_{2n+1}(us)Z_n(u^{-2})\,du$$

which is obtained from (54) by Hankel's inversion formula, preceded by the change of variables $x = us$, where $s^2t = 1$, is convergent but its value, instead of being $s^{-1}M(s^2)$ is $s^{-1}M_n(s^2)$ where $M_n(s^2)$ is the polynomial of degree $2n$ obtained by truncating the series for $M(s^2)$. We may thus write

$$M(x) = \lim_{n\to\infty} \int_0^\infty Z(xt^{-2})J_{2n+1}(t)\,dt. \tag{55}$$

It may be recalled that Forsyth's analysis shows that the roots of the function $f(z, 0, 0)$ are all real and negative, the numerically smallest roots being $(-1.058)^3$, $(-2.234)^3$, $(-3.437)^3$, $(-4.653)^3$ approximately, while for large values of n, the n^{th} root z_n is given approximately by the formula

$$z_n = \frac{\pi^3\sqrt{3}}{6561}\,(1 - 6n)^3. \tag{56}$$

A further property of the function M is expressed by the equation

$$\int_0^\infty M(x^2t)e^{-t}t^{\nu-1}\,dt = \frac{2}{\pi}\sin(\nu\pi)\int_0^{\frac{1}{2}\pi} J_0(x\sin\theta)\,(\tan\theta)^{2\nu-1}\,d\theta \tag{57}$$

$$0 < \nu < 1.$$

California Institute of Technology.

Annals of Mathematics
Vol. 38, No. 2, April, 1937

ON THE ORDER AND TYPE OF INTEGRAL FUNCTIONS BOUNDED AT A SEQUENCE OF POINTS

By V. Ganapathy Iyer

(Received September 1, 1936)

1. Introduction. Let $f(z)$ be an integral function. Let

$$M(r, f) = \max_{|z| \leq r}: |f(z)|.$$

The order ρ and the type $k(f)$ of $f(z)$ are defined by the relations

$$\rho = \overline{\lim_{r \to \infty}} \frac{\log \log M(r, f)}{\log r}; \qquad k(f) = \overline{\lim_{r \to \infty}} \frac{\log M(r, f)}{r^\rho}.$$

The following theorem has been proved by Miss Cartwright:[1]

Theorem (A): *Let $f(z)$ be an integral function of order one and type $k(f) < \pi$. Let $f(\pm n) = O(1)$, $n = 1, 2, 3 \ldots$. Then $f(z)$ is bounded along the whole of the real axis.*

1.1. The following result is an easy deduction from theorem (A):

Theorem (1): *Let $f(z)$ be an integral function of order one and type less than π. Let $f(\pm n) = O(1)$, $f(\pm in) = O(1)$. Then $f(z)$ will reduce to a constant.*

Proof: From theorem (A), we deduce that $f(z)$ is bounded along the whole of the real and imaginary axes. By applying a classical theorem of Phragmen and Lindelof,[2] we conclude that $f(z)$ is bounded in the whole plane and therefore is a constant.

1.2. Theorem (1) states that the type of an integral function of order one bounded at the points $z = \pm n, z = \pm in$ cannot be less than π unless it reduces to a constant. This is a special solution of the general problem of determining the order and type of an integral function taking a set of bounded values at a prescribed sequence of points. The author has considered elsewhere[3] some cases of this general problem where the points at which the function of order one is supposed to be bounded are distributed along a single line. The results are correspondingly less precise than those obtained in this paper where the function is supposed to be bounded at points lying on two lines as in theorem (1). Similar results are obtained when $0 < \rho < 1$, by supposing that the function is bounded

[1] "On certain integral functions of order one," Quar. Jour. Math., Oxford Series, Vol. 7, No. 25, (1936), 46–55.

[2] See, G. Valiron, "Lectures on Integral Functions," p. 125.

[3] "On integral functions of order one and finite type," Jour. Ind. Math. Soc., New Series, Vol. 2, No. 1. We shall refer to this paper as I in the sequel.

at sequences of points lying on both halves of a line instead of on one half only as is the case considered in another paper[4] by the author. The case $\rho > 1$ is not treated since the analogous results could be obtained by the process employed in §§3–3.5 of II. Theorem (A) itself does not seem capable of generalisation since it depends essentially on the fact that $\sin \pi z$, the canonical product for the sequence $\{\pm n\}$, is bounded along the whole of the real axis. This is not true always even when we restrict ourselves to sub-sequences of integers. For instance, the canonical product for $\{\pm\lambda_n\}$ where $\{\lambda_n\}$ is composed of the odd integers along with the even squares, is

$$\cos \tfrac{1}{2}\pi z \cdot \frac{\sin \frac{1}{2}\pi \sqrt{z}}{\frac{1}{2}\pi \sqrt{z}} \cdot \frac{\sin \frac{1}{2}\pi i \sqrt{z}}{\frac{1}{2}\pi i \sqrt{z}}$$

which is not bounded along the real axis.

1.3. Let $\{z_n\}$ be a distinct sequence of complex numbers such that $|z_n| \to \infty$. We shall speak of its exponent of convergence ρ as its order. It is supposed that $0 < \rho < \infty$. Let $\sigma(z)$ be the canonical product with simple zeroes at $z = z_n$. It is well-known that the order of $\sigma(z)$ is ρ. We write, for shortness,

$$P[f, \sigma] = \sum_{n=1}^{\infty} \frac{f(z_n)}{\sigma'(z_n)} \frac{1}{z - z_n}, \tag{1}$$

where $f(z)$ is any integral function and the series on the right side of (1) converges absolutely and uniformly for all z except at $z = z_n$; this happens to be the case in all applications considered below. $P[f, \sigma]$ represents a meromorphic function with simple poles at $z = z_n$ and having the same principal part as the function $f(z)/\sigma(z)$. Let each term in (1) be replaced by its modulus. The resulting expression is denoted by $\bar{P}[f, \sigma]$.

1.4. Let $0 < \lambda_1 < \lambda_2 \cdots < \lambda_n < \cdots$ be a sequence tending to infinity whose order is one. Let $n(r)$ denote the number of these λ_n not exceeding r. The sequence $\{\lambda_n\}$ is said to be measurable and of density D when $\lim_{r\to\infty} n(r)/r = D$. It is easy to see that $\sum 1/\lambda_n^2$ converges in this case. Let

$$A(z) = \prod_{n=1}^{\infty} \left(1 - \frac{z^2}{\lambda_n^2}\right). \tag{2}$$

Let δ be the index of condensation[5] of $\{\lambda_n\}$. The following results are known[5] and will be required in the sequel:

---

[4] "On integral functions of finite order bounded at a sequence of points," Ibid., No. 2. This paper will be referred to as II in the sequel.

[5] For the definition of the index of condensation, see, Bernstein, "Séries de Dirichlet," Borel Tracts, 1933. When $\{\lambda_n\}$ is measurable the index is given by the relation (4). When $\underline{\lim}_{n\to\infty}(\lambda_{n+1} - \lambda_n) > 0$ the index is zero. For the properties quoted, see pp. 279–293 of Bernstein's treatise.

THEOREM (B): *Let* $z = re^{i\psi}$. *The following relations are valid:*

$$\lim_{r\to\infty} \frac{\log |A(re^{i\psi})|}{r} = \pi D\, |\sin\psi|$$

$$\lim_{r\to\infty} \frac{\log M(r, A)}{r} = \pi D \tag{3}$$

where ψ *is not a multiple of* π. *Also*

$$\delta = \overline{\lim_{n\to\infty}} \frac{1}{\lambda_n} \log \frac{1}{|A'(\lambda_n)|}. \tag{4}$$

1.5. Let $\{\lambda_n\}$ be a sequence of order ρ, $0 < \rho < 1$, and let $\lim_{r\to\infty} n(r)/r^\rho = D$. We say that $\{\lambda_n\}$ is measurable of order ρ and density D. Let

$$A(z) = \prod_{n=1}^{\infty} \left(1 - \frac{z}{\lambda_n}\right). \tag{5}$$

Here, we define the index of condensation of $\{\lambda_n\}$ by the relation

$$\delta = \overline{\lim_{n\to\infty}} \frac{1}{\lambda_n^\rho} \log \frac{1}{|A'(\lambda_n)|}. \tag{6}$$

It may be noted, in passing, that the index as defined by (6) can be negative which cannot happen for the index as understood in §1.4, for $\rho = 1$ (see §5 below).

The following properties of $A(z)$ given by (5) are known:[6]

THEOREM (C): $A(z)$ *given by* (5) *satisfies the relations*:

$$\lim_{r\to\infty} \frac{\log M(r, A)}{r^\rho} = \pi D \operatorname{cosec} \pi^\rho$$

$$\lim_{r\to\infty} \frac{\log |A(re^{i\psi})|}{r^\rho} = \pi D \operatorname{cosec} \pi^\rho \cos \rho(\pi - |\psi|) \tag{7}$$

where ψ *is not a multiple of* 2π.

2. In this section, we are concerned with functions of order one. We prove the following

THEOREM (2): *Let* $\{\lambda_n\}$ *and* $\{\mu_n\}$ *be two measurable sequences of order one and densities* D_λ *and* D_μ *respectively. Let their indices of condensation be zero. Let* $f(z)$ *be an integral function of order one such that* $f(\pm\lambda_n) = O(1)$, $f(\pm i\mu_n) = O(1)$. *Then, either* $k(f) \geqq \pi D = \min: (\pi D_\lambda, \pi D_\mu)$ *or else* $f(z)$ *reduces to a constant.*

PROOF: If $D = 0$ the result to be proved is trivial. So we suppose $D > 0$. We assume $k(f) < \pi D$ and show that $f(z)$ is a constant. Let $A(z)$ and $B(z)$ be the canonical products for $\{\pm\lambda_n\}$ and $\{\pm i\mu_n\}$ respectively. Let us write

[6] See, Paley and Wiener, "Fourier Transforms in the Complex Domain," Amer. Math. Soc. Collo. Publications, XIX, p. 79.

$\lambda_n = -\lambda_{-n}$, $\mu_n = -\mu_{-n}$, for convenience. Let $F_m(z) = z^m f(z)$, $m = 0, 1, 2, \cdots$. We have

$$(8)\quad P[F_m, AB] = \sum_{n=-\infty}^{\infty}{}' \frac{F_m(\lambda_n)}{A'(\lambda_n)B(\lambda_n)} \frac{1}{z-\lambda_n} + \sum_{n=-\infty}^{\infty}{}' \frac{F_m(i\mu_n)}{A(i\mu_n)B'(i\mu_n)} \frac{1}{z-i\mu_n},$$

where the accent, as usual, indicates that $n = 0$ is to be omitted. The indices of condensation being zero, we get by (3) and (4)

$$(9)\quad \begin{aligned} &\overline{\lim_{|n|\to\infty}} \frac{1}{|\lambda_n|} \log \left| \frac{F_m(\lambda_n)}{A'(\lambda_n)B(\lambda_n)} \right| \leqq -\pi D_\mu \leqq -\pi D. \\ &\overline{\lim_{|n|\to\infty}} \frac{1}{|\mu_n|} \log \left| \frac{F_m(i\mu_n)}{A(i\mu_n)B'(i\mu_n)} \right| \leqq -\pi D_\lambda \leqq -\pi D. \end{aligned}$$

We see from (9) that (8) converges absolutely and uniformly for all z except at the zeroes of $A(z)\,B(z)$. Put

$$(10)\quad H_m(z) = \frac{F_m(z)}{A(z)B(z)} - P[F_m, AB].$$

By using the methods of I, §§4–4.3, we conclude that $H_m(z)$ is an integral function of order one. Also it is easy to see, by using (9), that on any line amp.$(z) = \theta$, not a multiple of $\frac{1}{2}\pi$,

$$(11)\quad \lim_{|z|\to\infty} P[F_m, AB] = 0.$$

On the other hand, on amp.$(z) = \theta$, not a multiple of $\frac{1}{2}\pi$, we get

$$(12)\quad \overline{\lim_{|z|\to\infty}} \frac{1}{|z|} \log \left| \frac{F_m(z)}{A(z)B(z)} \right| \leqq k(f) - \pi[D_\lambda |\sin\theta| + D_\mu |\cos\theta|]$$

in virtue of (3). Now $D_\lambda |\sin\theta| + D_\mu |\cos\theta|$ attains the maximum $\sqrt{(D_\lambda^2 + D_\mu^2)}$ on the four lines $\theta = \pm\alpha$, $\theta = \pm(\pi - \alpha)$, where $0 < \alpha < \frac{1}{2}\pi$. Hence on these lines

$$(13)\quad \lim_{|z|\to\infty} \left| \frac{F_m(z)}{A(z)B(z)} \right| = 0$$

since $k(f) < \pi D < \pi \cdot \sqrt{(D_\lambda^2 + D_\mu^2)}$. Combining (11) and (13), we find

$$\lim_{|z|\to\infty} H_m(z) = 0$$

along these four lines. Using the Phragmen-Lindelof theorem mentioned in §1.1, we conclude that $H_m(z) \equiv 0$ so that

$$(14)\quad \frac{z^m f(z)}{A(z)B(z)} = P[z^m f, AB]$$

Now, let $g(z) = c_0 + c_1 z + c_2 z^2 + \cdots$ be any integral function of order one and type $k(g) < \pi D$. Then

$$\sum |c_m| r^m \leqq e^{\beta r}$$

where $\beta < \pi D$. Hence, in virtue of (9), the double series

$$\sum_{(m,n)} \bar{P}[c_m z^m f, AB]$$

converges. So we get by (14),

$$\frac{f(z)g(z)}{A(z)B(z)} = P[fg, AB]. \tag{15}$$

In particular, we can take $g(z) = z^m f(z)$, $m = 0, 1, \cdots$, so that

$$\frac{z^m [f(z)]^2}{A(z)B(z)} = P[z^m f^2, AB]. \tag{16}$$

Starting from (16) we can repeat the argument and conclude that

$$\frac{[f(z)]^p}{A(z)B(z)} = P[f^p, AB] \tag{17}$$

for all $p = 0, 1, 2, \cdots$. Since $f(z)$ is bounded at the zeroes of $A(z)B(z)$, it follows from (9) that the double series

$$\sum_{(p,n)} \bar{P}\left[\frac{f^p}{p!}, AB\right]$$

converges so that (17) gives

$$\frac{e^{f(z)}}{A(z)B(z)} = P[e^f, AB]. \tag{18}$$

Since $e^{f(z)} = O(1)$ at the zeroes of $A(z)B(z)$, the methods of I §§4–4.3, enable us to conclude that

$$e^{f(z)} = A(z)B(z)\; P[e^f, AB]$$

is an integral function of order one. This shows that $f(z)$ is a polynomial of degree one at most. But $f(z)$ is bounded at the zeroes of $A(z)B(z)$. Hence $f(z)$ reduces to a constant.

2.1. By putting a more stringent hypothesis on the values of $f(z)$ at the zeroes of $A(z)B(z)$, we can obtain more precise results. We, first, prove

THEOREM (3): *Let $\{\lambda_n\}$ and $\{\mu_n\}$ be two measurable sequences of order one and densities D_λ and D_μ. Let their indices of condensation be δ_λ and δ_μ. Let $f(z)$ be a function of order one such that*

$$\begin{aligned} &\overline{\lim_{n\to\infty}} \frac{1}{\lambda_n} \log |f(\pm\lambda_n)| = -d_\lambda < 0. \\ &\overline{\lim_{n\to\infty}} \frac{1}{\mu_n} \log |f(\pm i\mu_n)| = -d_\mu < 0. \end{aligned} \tag{19}$$

Let $\pi D = \min: [\pi\sqrt{(D_\lambda^2 + D_\mu^2)}, \pi D_\lambda + d_\mu - \delta_\mu, \pi D_\mu + d_\lambda - \delta_\lambda]$. *Then* $k(f) \geqq \pi D$ *or else $f(z)$ is identically zero.*

Proof: In virtue of (19) we get in the notation of §2,

$$\overline{\lim_{|n|\to\infty}} \frac{1}{|\lambda_n|} \log \left| \frac{F_m(\lambda_n)}{A'(\lambda_n)B(\lambda_n)} \right| \leqq -(\pi D_\mu + d_\lambda - \delta_\lambda). \tag{20}$$
$$\overline{\lim_{|n|\to\infty}} \frac{1}{|\mu_n|} \log \left| \frac{F_m(i\mu_n)}{A(i\mu_n)B'(i\mu_n)} \right| \leqq -(\pi D_\lambda + d_\mu - \delta_\mu).$$

The case $D = 0$ is trivial. So we suppose $D > 0$. Then (20) shows that $P[F_m, AB]$ is absolutely and uniformly convergent except at the zeroes of $A(z)\,B(z)$ and we can repeat the argument of §2 to show that

$$\frac{F_m(z)}{A(z)B(z)} = \frac{z^m f(z)}{A(z)B(z)} = P[z^m f, AB] \tag{21}$$

when $k(f) < \pi D$, if we remember that (13) is valid if only $k(f) < \pi\sqrt{(D_\lambda^2 + D_\mu^2)}$. If $g(z)$ be a function of order one and type $k(g) < \pi D$, the relation (15) and hence all the relations that follow hold good so that we conclude that $f(z)$ is a constant. By (19) this constant is zero.

2.2. An interesting case is when $d_\lambda = d_\mu = \infty$. Here we can suppose that δ_λ and δ_μ are merely finite. So we get

Theorem (4): *Let $\{\lambda_n\}$ and $\{\mu_n\}$ be two measurable sequences of order one and densities D_λ and D_μ. Let their indices of condensation be finite. Let*

$$\lim_{n\to\infty} \frac{1}{\lambda_n} \log |f(\pm\lambda_n)| = \lim_{n\to\infty} \frac{1}{\mu_n} \log |f(\pm i\mu_n)| = -\infty,$$

where $f(z)$ is an integral function of order one. Then $k(f) \geqq \pi\sqrt{(D_\lambda^2 + D_\mu^2)}$ or else $f(z)$ is identically zero.

2.3. Taking $\lambda_n = \mu_n = n$, we have $\delta_\lambda = \delta_\mu = 0$ and so we can state

Theorem (4-a): *Let $f(z)$ be an integral function of order one such that*

$$\lim_{n\to\infty} \frac{1}{n} \log |f(\pm n)| = \lim_{n\to\infty} \frac{1}{n} \log |f(\pm in)| = -\infty.$$

Then the type of $f(z)$ cannot be less than $\pi\sqrt{2}$ unless $f(z)$ is identically zero. In particular, this is true when $|f(\pm n)| \leqq e^{-\alpha n \log n}$ and $|f(\pm in)| \leqq e^{-\beta n \log n}$ where α and β are positive.

2.4. **Remark**: Theorem (4-a) shows that the result of theorem (4) is a best possible one since the function $(\sin \pi z \sinh \pi z)/z^2$ which vanishes at $z = \pm n$, $z = \pm in$, $n = 1, 2, \cdots$ satisfies the condition of theorem (4) and has precisely $\pi\sqrt{2}$ for its type.

2.5. The significance of these theorems can be best realised by taking some concrete examples. Let $\{u_{\pm n}\}$ and $\{v_{\pm n}\}$, $n = 1, 2, \cdots$ be two bounded sequences, not all the terms being equal to a fixed number. Then we can assert that *the function*

$$\sin \pi z \sinh \pi z \left[\sum_{n=-\infty}^{\infty}{}' \frac{(-1)^n u_n}{\sinh \pi n} \frac{1}{z-n} + \sum_{n=-\infty}^{\infty}{}' \frac{(-1)^n v_n}{i \sinh \pi n} \frac{1}{z - in} \right]$$

is of order one and type $\geqq \pi$. The accent, as usual, indicates that $n = 0$ is to be omitted. If $u_{\pm n} = v_{\pm n} = e^{-n \log n}$, $n = 1, 2, \cdots$ we can assert that *the function*

$$\sin \pi z \sinh \pi z \left[\sum_{n=-\infty}^{\infty}{}' \frac{(-1)^n e^{-|n| \log |n|}}{\sinh \pi n} \frac{1}{z - n} + \sum_{n=-\infty}^{\infty}{}' \frac{(-1)^n e^{-|n| \log |n|}}{i \sinh \pi n} \frac{1}{z - in} \right]$$

is of order one and type $\geqq \pi \sqrt{2}$.

3. We treat the cases $0 < \rho \leqq \frac{1}{2}$ and $\frac{1}{2} < \rho < 1$ separately. We first prove

THEOREM (5): *Let* $\{\lambda_n\}$ *and* $\{\mu_n\}$ *be two measurable sequences of order* ρ, $0 < \rho \leqq \frac{1}{2}$, *and densities* D_λ *and* D_μ. *Let their indices of condensation be zero or negative. Let* $f(z)$ *be a function of order* ρ *such that* $f(\lambda_n) = O(1)$, $f(-\mu_n) = O(1)$ *then* $k(f) \geqq \pi D \operatorname{cosec} \pi\rho$ *where* $D = \min : (D_\lambda, D_\mu)$ *or else* $f(z)$ *is a constant.*

PROOF: We suppose $D > 0$ since $D = 0$ is trivial. Let

$$A(z) = \prod_{n=1}^{\infty} \left(1 - \frac{z}{\lambda_n}\right) \tag{22}$$
$$B(z) = \prod_{n=1}^{\infty} \left(1 + \frac{z}{\mu_n}\right).$$

We suppose $k(f) < \pi D \operatorname{cosec} \pi\rho$ and show that $f(z)$ reduces to a constant. Let, as before, $F_m(z) = z^m f(z)$. We have

$$P[F_m, AB] = \left[\sum_{n=1}^{\infty} \frac{F_m(\lambda_n)}{A'(\lambda_n)B(\lambda_n)} \frac{1}{z - \lambda_n} + \sum_{n=1}^{\infty} \frac{F_m(-\mu_n)}{A(-\mu_n)B'(-\mu_n)} \frac{1}{z + \mu_n} \right]. \tag{23}$$

Now, by (6) and (7),

$$\overline{\lim_{n \to \infty}} \frac{1}{\lambda_n^\rho} \log \left| \frac{F_m(\lambda_n)}{A'(\lambda_n)B(\lambda_n)} \right| \leqq -\pi D_\mu \operatorname{cosec} \pi\rho \leqq -\pi D \operatorname{cosec} \pi\rho \tag{24}$$
$$\overline{\lim_{n \to \infty}} \frac{1}{\mu_n^\rho} \log \left| \frac{F_m(-\mu_n)}{A(-\mu_n)B'(-\mu_n)} \right| \leqq -\pi D_\lambda \operatorname{cosec} \pi\rho \leqq -\pi D \operatorname{cosec} \pi\rho$$

since the indices of condensation are zero. Hence $P[F_m, AB]$ converges absolutely and uniformly except at the zeroes of $A(z)B(z)$. Let

$$H_m(z) = \frac{F_m(z)}{A(z)B(z)} - P[F_m, AB]. \tag{25}$$

By using the methods of II, §§2.31–2.32, we conclude that $H_m(z)$ is an integral function of order ρ. On amp. $(z) = \theta$, not a multiple of π, we have, from (23) and (7),

$$\lim_{|z| \to \infty} P[F_m, AB] = 0 \tag{26}$$

and

$$\overline{\lim_{|z| \to \infty}} \frac{1}{|z|^\rho} \log \left| \frac{F_m(z)}{A(z)B(z)} \right| \leqq K(f) - \pi \operatorname{cosec} \pi\rho [D_\lambda \cos \rho(\pi - |\theta|) + D_\mu \cos \rho|\theta|]. \tag{27}$$

The expression $D_\lambda \cos \rho(\pi - |\theta|) + D_\mu \cos \rho |\theta|$ attains the maximum $\sqrt{(D_\lambda^2 + D_\mu^2 + 2D_\lambda D_\mu \cos \pi\rho)}$ on a line $\theta = \alpha, 0 < \alpha < \pi$. Since $k(f) < \pi D$ cosec $\pi\rho$, (27) shows that on $\theta = \alpha$

$$\lim_{|z|\to\infty} \frac{F_m(z)}{A(z)B(z)} = 0. \tag{28}$$

If $0 < \rho < \frac{1}{2}$, we conclude from (26), (28) and the Phragmen-Lindelof theorem that $H_m(z) \equiv 0$. If $\rho = \frac{1}{2}$, (28) is true on lines $\theta = \alpha \pm \eta$, where $\eta > 0$ is sufficiently small. We can again apply the Phragmen-Lindelof theorem and deduce that $H_m(z) \equiv 0$. So we get

$$\frac{z^m f(z)}{A(z)B(z)} = P[z^m f, AB]. \tag{29}$$

With (29) as the starting point we can repeat the argument of §2 and show that $f(z)$ reduces to a constant.

3.1. The analogue of theorem (4) runs as follows and can be proved similarly:

THEOREM (6): *Let $\{\lambda_n\}$ and $\{\mu_n\}$ be two measurable sequences of order ρ, $0 < \rho \leqq \frac{1}{2}$ and densities D_λ and D_μ. Let their indices of condensation be finite. Let $f(z)$ be an integral function of order ρ such that*

$$\lim_{n\to\infty} \frac{1}{\lambda_n^\rho} \log |f(\lambda_n)| = \lim_{n\to\infty} \frac{1}{\mu_n^\rho} \log |f(-\mu_n)| = -\infty.$$

Then $k(f) \geqq \pi$ cosec $\pi\rho \sqrt{(D_\lambda^2 + D_\mu^2 + 2D_\lambda D_\mu \cos \pi\rho)}$ or else $f(z)$ is identically zero.

4. We now consider the case $\frac{1}{2} < \rho < 1$. Here we prove

THEOREM (7): *Let $\{\lambda_n\}$ and $\{\mu_n\}$ be as in theorem (5), the order ρ being such that $\frac{1}{2} < \rho < 1$. Let $f(z)$ be a function of order ρ, such that $f(\lambda_n) = O(1)$, $f(-\mu_n) = O(1)$. Then $k(f) \geqq \pi D$, where D is the greater of the two numbers* min : $(D_\lambda, D_\mu$ cosec $\pi\rho)$, *and* min : $(D_\mu, D_\lambda$ cosec $\pi\rho)$, *or else $f(z)$ reduces to a constant.*

PROOF: Suppose $k(f) < \pi D$. In the notation of §3, we find that the relations (22) to (26) are valid. It is in the estimation of

$$D_\lambda \cos \rho(\pi - |\theta|) + D_\mu \cos \rho|\theta|$$

that difficulties are introduced owing to the fact that $\cos \pi\rho$ is negative. But in any case, the expression π cosec $\pi\rho\, [D_\lambda \cos \rho(\pi - |\theta|) + D_\mu \cos \rho |\theta|]$ takes the value πD_λ for $\theta = \pm \frac{1}{2}\pi/\rho$ and πD_μ for $\theta = \pm (\pi - \frac{1}{2}\pi/\rho)$. If we denote by α the angle $\frac{1}{2}\pi/\rho$ or $\pi - \frac{1}{2}\pi/\rho$ according as D_λ or D_μ is the greater, we find that (28) is valid on the lines $\theta = \pm \alpha \pm \eta$ where $\eta > 0$ is sufficiently small. We can now apply the Pragmen-Lindelof theorem and conclude that (29) is true, from which we deduce that $f(z)$ is a constant.

4.1. Similarly we prove

THEOREM (8): *When $\{\lambda_n\}$ and $\{\mu_n\}$ are supposed to have merely finite indices of condensation, the conclusion of theorem (7) is true if $f(z)$ is such that*

$$\lim_{n\to\infty} \frac{1}{\lambda_n} \log |f(\lambda_n)| = \lim_{n\to\infty} \frac{1}{\mu_n^\rho} \log |f(-\mu_n)| = -\infty.$$

4.2. When the densities are equal, $D_\lambda = D_\mu = D$, say, the expression

$$D[\cos\rho(\pi - |\theta|) + \cos\rho|\theta|]$$

attains its maximum value $2D\cos\frac{1}{2}\pi\rho$ on the four lines $\theta = \pm\alpha$, $\theta = \pm(\pi - \alpha)$ where $0 < \infty < \pi$. Therefore if $\beta(\rho) = \min:(\operatorname{cosec}\pi\rho, \operatorname{cosec}\frac{1}{2}\pi\rho)$, we can state

THEOREM (9): *Let $\{\lambda_n\}$ and $\{\mu_n\}$ be two measurable sequences of order ρ, $\frac{1}{2} < \rho < 1$ and equal density D. Let their indices of condensation be zero or negative. Let $f(z)$ be of order ρ such that $f(\lambda_n) = O(1)$, $f(-\mu_n) = O(1)$. Then $k(f) \geqq \pi D\,\beta(\rho)$ or else $f(z)$ is a constant. If the indices are supposed merely finite, the conclusion remains true provided*

$$\lim_{n\to\infty} \frac{1}{\lambda_n^\sigma} \log |f(\lambda_n)| = \lim_{n\to\infty} \frac{1}{\mu_n^\rho} \log |f(-\mu_n)| = -\infty.$$

4.3. The difference in the results for $0 < \rho \leqq \frac{1}{2}$ and $\frac{1}{2} < \rho < 1$ arises from the fact that $\cos\pi\rho \geqq 0$ in one case and $\cos\pi\rho < 0$ in the other case. The maximum $\sqrt{(D_\lambda^2 + D_\mu^2 + 2D_\lambda D_\mu \cos\pi\rho)}$ of $D_\lambda\cos\rho(\pi - |\theta|) + D_\mu\cos\rho|\theta|$ can be equal to $D_\lambda\sin\pi\rho$ if $D_\mu + D_\lambda\cos\pi\rho = 0$ which can always be verified by adjusting the ratio D_μ/D_λ. If also $\frac{3}{4} < \rho < 1$, the conclusion of theorem (7) is a best possible one, since in this case the the function $A(z)B(z)$ itself has the type πD_λ. Therefore when $\frac{1}{2} < \rho < 1$ the ratio of the densities play a large part in the statement of the conclusions of theorem (7) and (8) but in no case can the type go below the limit given in these theorems.

5. As an illustration, let $\lambda_n = \mu_n = n^{1/\rho}$, $0 < \rho < 1$. It is shown in II, §4.2, that

$$\lim_{n\to\infty} \frac{1}{n} \log \left| \frac{1}{A'(n^{1/\rho})} \right| = -\pi\cot\pi\rho.$$

Therefore we can state the following theorems:

THEOREM (5-a): *If $f(z)$ is a function of order ρ, $0 < \rho \leqq \frac{1}{2}$, and $f(\pm n^{1/\rho}) = O(1)$, then $k(f) \geqq \pi\operatorname{cosec}\pi\rho$ unless $f(z)$ is a constant.*

THEOREM (6-a): *If $f(z)$ is of order ρ, $0 < \rho \leqq \frac{1}{2}$, and*

$$\lim_{n\to\infty} (1/n)\log|f(\pm n^{1/\rho})| = -\infty,$$

the type $k(f) \geqq \pi\operatorname{cosec}\frac{1}{2}\pi\rho$, or else $f(z)$ is identically zero.

THEOREM (9-a): *If $f(z)$ is of order ρ, $\frac{1}{2} < \rho < 1$ and*

$$\lim_{n\to\infty} (1/n) \log | f(\pm n^{1/\rho}) | = -\infty,$$

then $k(f) \geqq \pi \min : (\operatorname{cosec} \pi\rho, \operatorname{cosec} \frac{1}{2}\pi\rho)$, or else $f(z) \equiv 0$.

5.1. Conclusion: In theorems (5)-(8), we can take into account the actual values of the indices of condensation especially when these are negative as in theorem (5-a) when $0 < \rho < \frac{1}{2}$. These results show that the type of a function bounded at a sequence of points of the type considered in this paper, is, in general, controlled by the densities and the indices of condensation of the sequences in question. In connection with theorem (5-a), it might be of interest to mention that when it is merely assumed that $f(n^{1/\rho}) = O(1)$ instead of $f(\pm n^{1/\rho}) = O(1)$ as in the theorem mentioned, it can be proved that $f(z)$ reduces to a constant if $k(f) < \pi \cot \pi\rho$.

MADRAS UNIVERSITY, INDIA.

Annals of Mathematics
Vol. 38, No. 2, April, 1937

MAXIMAL CONVERGENCE OF SEQUENCES OF HARMONIC POLYNOMIALS

By J. L. Walsh

(Received August 19, 1936)

1. **Introduction.** The concept of maximal convergence of sequences of polynomials in the complex variable z exerts a unifying influence in the entire theory of approximation to an analytic function by polynomials. The property of maximal convergence is shared by all sequences of polynomials of best approximation as measured in a large variety of ways, and is shared likewise by many sequences of polynomials defined by interpolation.

The general theory of approximation of harmonic functions in two variables by harmonic polynomials is analogous to the theory of approximation of analytic functions by polynomials in z, and has received some attention in recent years.[1] It is to be expected that the concept of maximal convergence would unify also the theory of approximation (especially best approximation) by harmonic polynomials, and it is the object of the present paper to show how that expectation can be fulfilled. As definite advances made in the present paper, we mention the following: introduction and study of the concept of maximal convergence of harmonic polynomials, in such detail that by analogy new results are suggested or proved even for maximal sequences of polynomials in z; detailed study of degree of approximation and best approximation by harmonic polynomials on a point set whose complement is multiply connected (hitherto untreated,—older methods are inapplicable); the obtaining of new significant results even for approximation in the sense of least squares as measured by integration over a single Jordan curve.

This new study of maximal convergence of sequences of harmonic polynomials is broadly analogous to the older study of maximal convergence of sequences of polynomials in z, and is here to be derived from the older study. Nevertheless there are many differences both of method and result, so that the new study is by no means a trivial or obvious application of the older study.

In the sequel terminology and notation not explicitly mentioned are uniform with the writer's recent book,[2] to which the reader is referred also for an account of the theory of maximal convergence of sequences of polynomials in the complex variable. References not otherwise indicated are to that book.

[1] See for instance a recent report, Walsh, Bull. Amer. Math. Soc., vol. 35 (1929), pp. 499-544.

[2] *Interpolation and Approximation by Rational Functions in the Complex Domain* (New York, 1935). The reader may refer to this book also for the history of approximation by polynomials in z.

Throughout the present paper we shall deal with the $z(= x + iy)$-plane. A function $u(x, y)$ is *harmonic at a point* if it is harmonic throughout a neighborhood of that point, and is *harmonic in a region* if there it is continuous together with its first and second partial derivatives, and satisfies Laplace's equation. Two functions $u(x, y)$ and $v(x, y)$ are *conjugate* in a region if there they are continuous together with their first and second partial derivatives, and if the functions satisfy the Cauchy-Riemann differential equations:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

If $u(x, y)$ and $v(x, y)$ are conjugate in a region, they are harmonic there. If $u(x, y)$ is given harmonic in a simply connected region, there exists a function $v(x, y)$ conjugate to $u(x, y)$ in that region; the function $v(x, y)$ is uniquely determined except for an additive constant, which is arbitrary. Whenever $u(x, y)$ and $v(x, y)$ are conjugate in a region, so also are $v(x, y)$ and $-u(x, y)$. If $u(x, y)$ is harmonic in a region, so are $\partial u/\partial x$ and $\partial u/\partial y$; the functions $\partial u/\partial x$ and $-\partial u/\partial y$ are conjugate there.

2 Lemmas on harmonic functions. It will be convenient to have for reference several easily proved and well known lemmas concerning harmonic functions.

LEMMA I. *Let S be a limited simply connected region, and let S_1 be an arbitrary closed subregion interior to S. There exists a number M' depending only on S and S_1, such that for every function $U(x, y)$ harmonic and in absolute value not greater than η in S we have*

$$(1) \qquad \left|\frac{\partial U}{\partial x}\right| \leqq M'\eta, \qquad \left|\frac{\partial U}{\partial y}\right| \leqq M'\eta, \qquad (x, y) \text{ in } S_1.$$

Let S_2 denote an arbitrary closed region bounded by an analytic Jordan curve, where S_2 lies interior to S but contains S_1 in its interior; such a region S_2 exists, as we see at once by mapping S onto the interior of a circle. Let $G(x, y; \alpha, \beta)$ denote Green's function for S_2 with pole in the point (x, y) and running coördinates (α, β). We have

$$(2) \qquad U(x, y) = \frac{1}{2\pi}\int U(\alpha, \beta)\,\frac{\partial G}{\partial \nu}\,ds_{(\alpha,\beta)},$$

where (x, y) is interior to S_2, ν is the interior normal, and the integral is taken over the boundary of S_2. By differentiation of (2) with respect to x and y respectively, and by the inequality $|U(x, y)| \leqq \eta$ valid on the boundary of S_2, we obtain inequalities (1).[3]

LEMMA II. *Let S_1 be an arbitrary closed region bounded by an analytic Jordan curve, and let (a, b) be a fixed point interior to S_1. Then there exists a number M''*

[3] Of course Lemma I can be proved also by covering S_1 with circles interior to S and by using Poisson's integral for the function $U(x, y)$ in each circle.

depending only on S_1 and (a, b) such that for a function $W(x, y)$ with continuous first and second partial derivatives in S_1 with the property

$$(3) \qquad \left| \frac{\partial W(x, y)}{\partial x} \right| \leqq \eta, \qquad \left| \frac{\partial W(x, y)}{\partial y} \right| \leqq \eta, \qquad (x, y) \text{ in } S_1,$$

we have also

$$(4) \qquad | W(x, y) - W(a, b) | \leqq M''\eta, \qquad (x, y) \text{ in } S_1 .$$

We have the well known relation

$$(5) \qquad W(x, y) - W(a, b) = \int_{(a,b)}^{(x,y)} \left(\frac{\partial W}{\partial x} dx + \frac{\partial W}{\partial y} dy \right), \qquad (x, y) \text{ in } S_1,$$

where the integral is taken over a path having no point exterior to S_1; this integral is then independent of the particular path chosen. There exist paths of the kind described from (a, b) to all points (x, y) of S_1, such that the paths are in length less than some number independent of (x, y).[4] Inequality (4) follows at once from equation (5) and inequalities (3).

LEMMA III. *Let S be a limited simply connected region, let S_1 be an arbitrary closed subregion interior to S bounded by an analytic Jordan curve, and let (a, b) be a fixed point interior to S_1. Then there exists a number M''' depending only on S, S_1, and (a, b) such that for every function $U(x, y)$ harmonic and in absolute value not greater than η in S we have*

$$| V(x, y) - V(a, b) | \leqq M'''\eta, \qquad (x, y) \text{ in } S_1 ,$$

where $V(x, y)$ is conjugate to $U(x, y)$ interior to S.

LEMMA I is applicable and yields inequalities (1). By the Cauchy-Riemann equations

$$\frac{\partial U}{\partial x} = \frac{\partial V}{\partial y}, \qquad \frac{\partial U}{\partial y} = -\frac{\partial V}{\partial x},$$

we thus have inequalities on the first partial derivatives of $V(x, y)$, and Lemma III follows from Lemma II.

LEMMA IV. *Let S be a limited simply connected region, let S_1 be an arbitrary closed subregion interior to S bounded by an analytic Jordan curve, and let (a, b) be a fixed point interior to S_1. Then there exists a number M^{iv} depending only on*

[4] It is obviously possible to choose such paths, for if we map the region S_1 onto the interior of the unit circle $\gamma : | z' | = 1$ by a transformation $z' = \Phi(z)$ so that (a, b) corresponds to the origin, we have

$$\int_{a+ib}^{x+iy} | dz | = \int_0^{z'} \left| \frac{dz'}{\Phi'(z)} \right| ;$$

the latter integral may be taken over a line segment (of length not greater than unity), and the function $1/| \Phi'(z) |$ is uniformly bounded on and within γ.

S, S_1, and (a, b) such that for every function $U(x, y)$ harmonic interior to S and for which we have

$$(6) \qquad \left|\frac{\partial U}{\partial x}\right| \leqq \eta, \qquad (x, y) \text{ in } S,$$

we have also

$$(7) \qquad \left|U(x, y) - U(a, b) - \frac{\partial U(a, b)}{\partial y}(y - b)\right| \leqq M^{iv}\eta, \qquad (x, y) \text{ in } S_1;$$

if for the function $U(x, y)$ harmonic interior to S we assume

$$\left|\frac{\partial U}{\partial y}\right| \leqq \eta, \qquad (x, y) \text{ in } S,$$

instead of (6) *then we have*

$$\left|U(x, y) - U(a, b) - \frac{\partial U(a, b)}{\partial x}(x - a)\right| \leqq M^{iv}\eta, \qquad (x, y) \text{ in } S_1.$$

Let S_2 denote a closed region interior to S, bounded by an analytic Jordan curve, and containing S_1 in its interior. The functions $\partial U/\partial x$ and $-\partial U/\partial y$ are conjugate in S, so by Lemma III inequality (6) implies

$$\left|\frac{\partial U(x, y)}{\partial y} - \frac{\partial U(a, b)}{\partial y}\right| \leqq M'''\eta, \qquad (x, y) \text{ in } S_2.$$

By inequality (6) and Lemma II we now have inequality (7). The proof is similar if we replace (6) by the corresponding inequality on $\partial U/\partial y$.

3. Degree of convergence, general sequences. The lemmas just established have immediate application to degree of convergence of sequences of harmonic functions.

THEOREM 1. *Let S be a limited simply connected region, and let S_1 be an arbitrary closed subregion interior to S bounded by an analytic Jordan curve. Let the functions $u_n(x, y)$ harmonic in S approach uniformly the function $u(x, y)$, necessarily harmonic in S:*

$$\lim_{n\to\infty} \eta_n = 0, \quad |u(x, y) - u_n(x, y)| \leqq \eta_n \quad \text{for } (x, y) \text{ in } S.$$

Then we have also

$$\left|\frac{\partial u(x, y)}{\partial x} - \frac{\partial u_n(x, y)}{\partial x}\right| \leqq M'\eta_n, \quad \left|\frac{\partial u(x, y)}{\partial y} - \frac{\partial u_n(x, y)}{\partial y}\right| \leqq M'\eta_n, \quad (x, y) \text{ in } S_1,$$

$$|f'(z) - f_n'(z)| \leqq \sqrt{2}\, M'\eta_n, \quad z \text{ in } S_1,$$

where $$f'(z) \equiv \frac{\partial u}{\partial x} - i\frac{\partial u}{\partial y}, \qquad f_n'(z) \equiv \frac{\partial u_n}{\partial x} - i\frac{\partial u_n}{\partial y},$$

and where M' is defined as in Lemma I.

Theorem 1 follows at once from Lemma I. The functions $f'(z)$ and $f'_n(z)$ are of course analytic in S, respective derivatives of the analytic functions

$$f(z) \equiv u(x, y) + iv(x, y), \qquad f_n(z) \equiv u_n(x, y) + iv_n(x, y),$$

where $v(x, y)$ and $v_n(x, y)$ are conjugate to $u(x, y)$ and $u_n(x, y)$ respectively in S.

THEOREM 2. *Under the hypothesis of Theorem 1 let (a, b) be a fixed point interior to S_1. Then we have*

$$|[v(x, y) - v(a, b)] - [v_n(x, y) - v_n(a, b)]| \leqq M'''\eta_n, \qquad (x, y) \text{ in } S_1,$$
$$|[f(z) - iv(a, b)] - [f_n(z) - iv_n(a, b)]| \leqq \sqrt{1 + M'''^2}\, \eta_n, \qquad z \text{ in } S_1,$$

where

$$f(z) \equiv u(x, y) + iv(x, y), \qquad f_n(z) \equiv u_n(x, y) + iv_n(x, y),$$

where $v(x, y)$ and $v_n(x, y)$ are conjugate to $u(x, y)$ and $u_n(x, y)$ respectively in S, and where M''' is defined as in Lemma III.

If the sequence $v_n(x, y)$ converges to $v(x, y)$ at even a single point of S_1, then $v_n(a, b)$ converges to $v(a, b)$, so $v_n(x, y)$ converges to $v(x, y)$ uniformly in S_1. If the sequence $f_n(z)$ converges to $f(z)$ at even a single point of S_1, then $v_n(a, b)$ converges to $v(a, b)$, so $f_n(z)$ converges to $f(z)$ uniformly in S_1.

Theorem 2 follows at once from Lemma III.

THEOREM 3. *Let S be a limited simply connected region, and let S_1 be an arbitrary closed subregion interior to S bounded by an analytic Jordan curve and containing a fixed point (a, b). Let the functions $u(x, y)$ and $u_n(x, y)$ be harmonic in S, and suppose we have*

$$\lim_{n\to\infty} \epsilon_n = 0, \qquad \left|\frac{\partial u}{\partial x} - \frac{\partial u_n}{\partial x}\right| \leqq \epsilon_n \quad \text{for } (x, y) \text{ in } S. \tag{8}$$

Then we have also

$$\left|[u(x, y) - u_n(x, y)] - [u(a, b) - u_n(a, b)] - \left[\frac{\partial u(a, b)}{\partial y} - \frac{\partial u_n(a, b)}{\partial y}\right](y - b)\right| \leqq M^{iv}\epsilon_n, \qquad (x, y) \text{ in } S_1,$$

$$\left|[v(x, y) - v_n(x, y)] - [v(a, b) - v_n(a, b)] - \left[\frac{\partial v(a, b)}{\partial x} - \frac{\partial v_n(a, b)}{\partial x}\right](x - a)\right| \leqq M^{iv}\epsilon_n, \qquad (x, y) \text{ in } S_1, \tag{9}$$

$$\left|[f(z) - f_n(z)] - [f(\alpha) - f_n(\alpha)] + i\left[\frac{\partial u(a, b)}{\partial y} - \frac{\partial u_n(a, b)}{\partial y}\right](z - \alpha)\right| \leqq \sqrt{2}\, M^{iv}\epsilon_n, \qquad z \text{ in } S_1,$$

where $f(z) \equiv u(x, y) + iv(x, y), \quad f_n(z) \equiv u_n(x, y) + iv_n(x, y), \quad \alpha = a + ib,$

where $v(x, y)$ and $v_n(x, y)$ are conjugate to $u(x, y)$ and $u_n(x, y)$ respectively in S, and where M^{iv} is defined as in Lemma IV.

The first of inequalities (9) follows at once from Lemma IV. The inequality that occurs in (8) can also be written

$$\left| \frac{\partial v}{\partial y} - \frac{\partial v_n}{\partial y} \right| \leqq \epsilon_n ,$$

so the second of inequalities (9) also follows from Lemma IV. The first and second of inequalities (9) imply the third of those inequalities.

Theorems 1, 2, and 3 concerning sequences of harmonic functions in general are now to be applied in particular to the study of sequences of harmonic polynomials.

A *polynomial in the complex variable z of degree n* is any function which can be written in the form $a_0 z^n + a_1 z^{n-1} + \cdots + a_n$. A *harmonic polynomial in x and y of degree n* is the real part of a polynomial in $z = x + iy$ of degree n. Thus if $p_n(x, y)$ is a harmonic polynomial, with $P_n(z) \equiv p_n(x, y) + iq_n(x, y)$, the function $q_n(x, y)$ conjugate to $p_n(x, y)$ (uniquely defined to within an additive constant) is also a harmonic polynomial of degree n, the real part of $-iP_n(z)$, a polynomial in z of degree n. Under these conditions the functions $\partial p_n/\partial x$ and $\partial p_n/\partial y$ are also harmonic polynomials, of degree $n - 1$, for we have

$$P_n'(z) \equiv \frac{\partial p_n}{\partial x} + i \frac{\partial q_n}{\partial x} \equiv \frac{\partial p_n}{\partial x} - i \frac{\partial p_n}{\partial y}.$$

4. **Degree of convergence, harmonic polynomials.** The following convention is usual [op. cit., p. 65] and is to be retained throughout the sequel:

NOTATION. *Let C be an arbitrary closed limited point set of the $z(= x + iy)$-plane whose complement K with respect to the extended plane is connected, and is regular in the sense that K possesses a Green's function $G(x, y)$ with pole at infinity. The function $w = \varphi(z) = e^{G+iH}$, where H is conjugate to G in K, maps K conformally but not necessarily uniformly onto the exterior of the circle $| w_\cdot | = 1$ so that the points at infinity in the two planes correspond to each other. Then C_R shall denote generically the locus $| \varphi(z) | = R > 1$ (or $G(x, y) = \log R > 0$) exterior to C; that is to say, C_R is the image of the circle $| w | = R > 1$ under the transformation $w = \varphi(z)$. The case $R = \infty$ is not excluded; the notation C_∞ refers to the point at infinity, and the "interior" of C_∞ is considered to be the entire plane with the omission of the point at infinity.*

The fundamental result concerning degree of convergence and overconvergence of sequences of harmonic polynomials is[5]

THEOREM 4. *Let C be a closed limited point set composed of the mutually exterior closed Jordan regions $J_1, J_2, \cdots, J_k$.*

---

[5] Theorem 4 has already been established for the case $k = 1$; Walsh, Bull. Amer. Math. Soc., vol. 33 (1927), pp. 591–598.

If the function $u(x, y)$ is defined merely on C, and if the harmonic polynomials $p_n(x, y)$ of respective degrees n satisfy the relation

$$(10)\quad \overline{\lim_{n\to\infty}}\, \mu_n^{1/n} = 1/\rho < 1, \qquad \mu_n = \max [| u(x, y) - p_n(x, y) |, (x, y) \text{ on } C],$$

then the sequence $p_n(x, y)$ converges throughout the interior of C_ρ, uniformly on any closed point set interior to C_ρ. Consequently the function $u(x, y)$ can be extended harmonically from C along paths interior to C_ρ so as to be single-valued and harmonic throughout the interior of C_ρ.

If the function $u(x, y)$ is single-valued and harmonic on C, there exists a number ρ finite or infinite such that $u(x, y)$ (or its harmonic extension) is single-valued and harmonic throughout the interior of C_ρ but not throughout the interior of any $C_{\rho'}$, $\rho' > \rho$. There exists a sequence of harmonic polynomials $p_n(x, y)$ of respective degrees n such that (10) is valid, but there exists no sequence of harmonic polynomials $p_n(x, y)$ of respective degrees n such that we have

$$(11)\quad \overline{\lim_{n\to\infty}}\, \mu_n^{1/n} = 1/\rho' < 1/\rho, \qquad \mu_n = \max [| u(x, y) - p_n(x, y) |, (x, y) \text{ on } C].$$

Let $u(x, y)$ be now defined on C, let (10) be valid, and let R be arbitrary, $1 < R < \rho$. Choose R_1 so that $R < R_1 < \rho$. From (10) we have

$$(12)\qquad | u(x, y) - p_n(x, y) | \leqq \frac{M}{R_1^n}, \qquad (x, y) \text{ on } C,$$

where M is suitably chosen. In the region J_j ($j = 1, 2, \cdots, k$) we choose a point (a_j, b_j), and define the functions

$$v_j(x, y) = \int_{(a_j,b_j)}^{(x,y)} \left(-\frac{\partial u}{\partial y}\, dx + \frac{\partial u}{\partial x}\, dy \right),$$

$$q_{nj}(x, y) = \int_{(a_j,b_j)}^{(x,y)} \left(-\frac{\partial p_n}{\partial y}\, dx + \frac{\partial p_n}{\partial x}\, dy \right),$$

conjugate respectively to $u(x, y)$ and $p_n(x, y)$ in J_j. Of course the functions $v_j(x, y)$ and $q_{nj}(x, y)$ depend effectively on j.

Let us now set

$$f_j(z) \equiv u(x, y) + iv_j(x, y), \qquad P_{nj}(z) \equiv p_n(x, y) + iq_{nj}(x, y),$$

so that $f_j(z)$ is analytic interior to J_j, and $P_{nj}(z)$ is a polynomial in z of degree n. We obtain at once from (12) by Theorem 1

$$(13)\qquad | f_j'(z) - P_{nj}'(z) | \leqq \frac{M_1}{R_1^n}, \qquad z \text{ on } J_j',$$

where J_j' is an arbitrary closed region interior to J_j bounded by an analytic Jordan curve, and where M_1 is suitably chosen.

If the region J_j' is suitably chosen, the closed region J_j lies [op. cit., §2.1,

Theorem 2] interior to $[J_j']_{R_1/R}$. For z on $[J_j']_{R_1/R}$ and consequently for z on J_j we have [op. cit., §4.7, Corollary to Theorem 8] from (13)

$$|f_j'(z) - P'_{nj}(z)| \leqq \frac{M_2}{R^n}, \qquad z \text{ on } J_j, \tag{14}$$

where M_2 is suitably chosen;[6] and M_2 need not depend on j.

We come now to the reason for the detailed study of the derivatives $f_j'(z)$ and $P'_{nj}(z)$ rather than the functions $f_j(z)$ and $P_{nj}(z)$ themselves; it is precisely in this respect that the proof of Theorem 4 for $k > 1$ differs essentially from the proof for $k = 1$: *the derivatives $f_j'(z)$ and $P'_{nj}(z)$ are independent of j.* Indeed, we may write

$$\begin{aligned} f_j'(z) &= \frac{\partial u}{\partial x} + i\,\frac{\partial v_j}{\partial x} = \frac{\partial u}{\partial x} - i\,\frac{\partial u}{\partial y}, \\ P'_{nj}(z) &= \frac{\partial p_n}{\partial x} + i\,\frac{\partial q_{nj}}{\partial x} = \frac{\partial p_n}{\partial x} - i\,\frac{\partial p_n}{\partial y}. \end{aligned} \tag{15}$$

The function $P'_{nj}(z)$ is a polynomial in z of degree $n - 1$, so inequality (14) valid on C for every $R < \rho$ implies [op. cit., §4.6, Theorem 6] the convergence of the sequence $P'_{nj}(z)$ to a function which may be denoted by $f_j'(z)$, throughout the interior of C_ρ, uniformly on any closed set interior to C_ρ.

It is appropriate to mention the independence of j not merely of $P'_{nj}(z)$ but also of $f_j'(z)$, for several points (a_j, b_j) may lie interior to a single Jordan region bounded by C_ρ or a part of C_ρ. We need to know that in such a region $f_j'(z)$ does not depend on j.

On any closed set J interior to C_ρ containing a point (a_j, b_j) we have from Theorem 3

$$\lim_{n\to\infty} p_n(x, y) = u(x, y), \qquad \text{uniformly in } J, \tag{16}$$

for $p_n(a_j, b_j)$ approaches $u(a_j, b_j)$, and by (14) the derivative $\partial p_n(a_j, b_j)/\partial y$ approaches $\partial u(a_j, b_j)/\partial y$. Each of the finite regions into which C_ρ separates the plane contains a point (a_j, b_j). The sequence $p_n(x, y)$ thus converges to a function which may be denoted by $u(x, y)$, throughout the interior of C_ρ, uniformly on any closed set interior to C_ρ. Hence the function $u(x, y)$ can be extended harmonically from C so as to be single-valued and harmonic throughout the interior of C_ρ. The first part of Theorem 4 is established.

---

[6] This reasoning is used many times in the book mentioned, for instance pp. 96–97. We may also prove and apply the following general theorem:

Let C be a closed limited Jordan region; suppose $\mu < 1$ is given, and we have on every closed subset C' interior to C

$$\overline{\lim_{n\to\infty}}\ \{\max\,[|f(z) - p_n(z)|,\ z \text{ on } C']\}^{1/n} \leqq \mu,$$

where $p_n(z)$ is a polynomial in z of degree n. Then we have likewise

$$\overline{\lim_{n\to\infty}}\ \{\max\,[|f(z) - p_n(z)|,\ z \text{ on } C]\}^{1/n} \leqq \mu.$$

Suppose now $u(x, y)$ is given single-valued and harmonic on C. The number ρ of Theorem 4 exists [compare the method of op. cit., p. 58]. Let $v(x, y)$ be a function conjugate to $u(x, y)$ interior to C_ρ; if the interior of C_ρ consists of several regions there need be no relation between the values of $u(x, y)$ or between the values of $v(x, y)$ in these various regions. The function $f(z) \equiv u(x, y) + iv(x, y)$ is single-valued and analytic throughout the interior of C_ρ. Consequently [op. cit., p. 79] there exist polynomials $P_n(z)$ in z of respective degrees n such that we have for an arbitrary $R < \rho$

$$|f(z) - P_n(z)| \leqq \frac{M}{R^n}, \qquad z \text{ on } C,$$

where M is suitably chosen. Therefore we have also

$$|u(x, y) - p_n(x, y)| \leqq \frac{M}{R^n}, \qquad (x, y) \text{ on } C,$$

where $P_n(z) \equiv p_n(x, y) + iq_n(x, y)$. This inequality yields

$$(17) \quad \overline{\lim_{n\to\infty}}\, \mu_n^{1/n} \leqq 1/\rho, \qquad \mu_n = \max\,[|u(x, y) - p_n(x, y)|, (x, y) \text{ on } C].$$

Either the inequality in (17) or equation (11) would imply (by the part of Theorem 4 already established) that $u(x, y)$ is single-valued and harmonic throughout the interior of some $C_{\rho'}$, $\rho' > \rho$, contrary to the definition of ρ. Then equation (10) and the impossibility of equation (11) are established, so the proof of Theorem 4 is complete.

A proposition less specific but somewhat simpler than Theorem 4 is: *Let C be a closed limited point set consisting of the mutually exterior closed Jordan regions $J_1, J_2, \cdots, J_k$. Let the function $u(x, y)$ be defined and single-valued on C. A necessary and sufficient condition that $u(x, y)$ be harmonic on C is that there exist harmonic polynomials $p_n(x, y)$ of respective degrees n such that equation* (10) *is fulfilled, where ρ is some number greater than unity.*

5. **Degree of convergence on C_σ.** A more explicit result than equation (16) can be obtained by inspection of the proof of (16). Let us choose an arbitrary C_σ, $\sigma < \rho$, having no multiple point. Let J be the closed interior of an arbitrary one of the analytic Jordan curves composing C_σ; then J contains in its interior at least one of the points (a_j, b_j). Inequality (14) implies [op. cit., p. 81, Corollary to Theorem 8]

$$|f_j'(z) - P_{nj}'(z)| \leqq \frac{M_3\sigma^n}{R^n}, \qquad z \text{ on or within } C_\sigma, \qquad \sigma < R,$$

where M_3 is suitably chosen. Consequently we may write for z in J

$$\left|\frac{\partial(u - p_n)}{\partial x}\right| \leqq \frac{M_3\sigma^n}{R^n}, \qquad \left|\frac{\partial(u - p_n)}{\partial y}\right| \leqq \frac{M_3\sigma^n}{R^n}.$$

As a particular case under (12) we have

$$|u(a_j, b_j) - p_n(a_j, b_j)| \leqq \frac{M}{R_1^n}, \qquad R_1 > R,$$

so by Lemma II we write the inequality

$$(18) \qquad |u(x, y) - p_n(x, y)| \leqq \frac{M_4 \sigma^n}{R^n},$$

valid in each of the closed regions J, thus (with the requisite modification in M_4 if necessary) valid throughout the closed interior of C_σ.

Inequality (18) has for convenience been established only when C_σ has no multiple points. If now C_σ has multiple points we may choose R', $R < R' < \rho$, and may choose σ', $\sigma < \sigma' < \sigma R'/R$, in such a manner that $C_{\sigma'}$ has no multiple points. By the proof of (18) we have

$$(19) \qquad |u(x, y) - p_n(x, y)| \leqq M_5\left(\frac{\sigma'}{R'}\right)^n \leqq M_5\left(\frac{\sigma}{R}\right)^n$$

throughout the closed interior of $C_{\sigma'}$, hence throughout the closed interior of C_σ. The arbitrariness of $R < \rho$ in (18) and (19) allows us now to formulate[7]

THEOREM 5. *Let C be a closed limited point set consisting of a finite number of mutually exterior closed Jordan regions. Let the function $u(x, y)$ be defined on C, and let the harmonic polynomials $p_n(x, y)$ of respective degrees n satisfy* (10). *Then we have also* $(1 < \sigma < \rho)$

$$(20) \quad \overline{\lim_{n\to\infty}}\, [\mu_n(\sigma)]^{1/n} \leqq \sigma/\rho, \qquad \mu_n(\sigma) = \max\,[|u(x, y) - p_n(x, y)|, (x, y) \text{ on } C_\sigma].$$

If under the conditions of Theorem 5 the function $u(x, y)$ is single-valued and harmonic throughout the interior of C_ρ but not throughout the interior of any $C_{\rho'}$, $\rho' > \rho$, then the inequality in (20) is impossible. This follows from the relation $[C_\sigma]_{\rho/\sigma} = C_\rho$ and from the last part of Theorem 4. That is to say, we have proved the

COROLLARY. *Under the hypothesis of Theorem 5 let $u(x, y)$ be single-valued and harmonic throughout the interior of C_ρ but not throughout the interior of any $C_{\rho'}$, $\rho' > \rho$. Then we have*

$$\overline{\lim_{n\to\infty}}\, [\mu_n(\sigma)]^{1/n} = \sigma/\rho.$$

---

[7] Both Theorem 5 and the first part of Theorem 4 are easily proved also from the following:

LEMMA. *Let C satisfy the hypothesis of Theorem 4. Let there be given R and R_1, $1 < R_1 < R$. There exists a number N depending only on C, R, and R_1, such that for every harmonic polynomial $\pi_n(x, y)$ of degree n of modulus not greater than η on C we have*

$$|\pi_n(x, y)| \leqq N\eta R^n, \qquad (x, y) \text{ on or within } C_{R_1}.$$

This lemma may be proved by Lemma I, by double application to the function $\partial\pi_n/\partial x - i\partial\pi_n/\partial y$ of the analogue [op. cit., p. 77] of the present lemma first to the point sets J_j' used in the proof of (14) and second to the set C, and finally by Lemma II.

It is convenient to formulate for reference several other results whose proofs are essentially contained in the proof of Theorems 4 and 5:

THEOREM 6. *Let C be a closed limited point set consisting of the mutually exterior closed Jordan regions $J_1, J_2, \cdots, J_k$. Let the function $u(x, y)$ be defined on C, and let the harmonic polynomials $p_n(x, y)$ of respective degrees n satisfy the relation*

$$(21)\quad \overline{\lim_{n\to\infty}}\,[\nu_n(J_j')]^{1/n} \leqq 1/\rho < 1, \qquad \nu_n(J_j') = \max\,[|\,u(x,y) - p_n(x,y)\,|,\,(x,y) \text{ in } J_j'],$$

for every closed Jordan region J_j' interior to J_j bounded by an analytic Jordan curve. Then we have also

$$(22)\quad \overline{\lim_{n\to\infty}}\,\mu_n^{1/n} \leqq 1/\rho, \quad \mu_n = \max\,[|\,u(x,y) - p_n(x,y)\,|,\,(x,y) \text{ on } C].$$

Inequality (21) instead of equation (10) is by Theorem 1 sufficient for the proof of (13), and hence is sufficient for the conclusion of Theorem 5. But we may obviously write

$$\mu_n \leqq \mu_n(\sigma),$$

whence from (20)

$$\overline{\lim_{n\to\infty}}\,\mu_n^{1/n} \leqq \sigma/\rho.$$

This last inequality holds for every σ, $1 < \sigma < \rho$, and hence yields (22).

THEOREM 7. *Let C be a closed limited point set consisting of the mutually exterior closed Jordan regions $J_1, J_2, \cdots, J_k$. Let the function $u(x, y)$ be defined on C, and let the harmonic polynomials $p_n(x, y)$ of respective degrees n satisfy the relation (22). Then we have also*

$$(23)\quad \overline{\lim_{n\to\infty}}\,\lambda_n'^{1/n} \leqq 1/\rho, \qquad \lambda_n' = \max\,[|\,f'(z) - P_n'(z)\,|,\,z \text{ on } C],$$

where $\quad f'(z) \equiv \partial u/\partial x - i\partial u/\partial y, \qquad P_n'(z) \equiv \partial p_n/\partial x - i\partial p_n/\partial y.$

In the case $k = 1$ we have

$$(24)\quad \overline{\lim_{n\to\infty}}\,\lambda_n^{1/n} \leqq 1/\rho, \qquad \lambda_n = \max\,[|\,f(z) - P_n(z)\,|,\,z \text{ on } C],$$

where $f(z) \equiv u + iv$, $P_n(z) \equiv p_n + iq_n$, and where $v(x, y)$ and $q_n(x, y)$ are conjugate respectively in C to $u(x, y)$ and $p_n(x, y)$, and vanish in a fixed point (a, b) interior to C.

Inequality (23) follows at once from Theorem 1 and Theorem 6.
Inequality (24) follows from Theorem 2 and Theorem 6.

6. Maximal convergence. Theorem 4 suggests of itself the formulation of the

DEFINITION. *Let C be a closed limited point set consisting of a finite number of mutually exterior closed Jordan regions $J_1, J_2, \cdots, J_k$. Let the function $u(x, y)$ (or its harmonic extension) be single-valued and harmonic throughout the interior of C_ρ but not throughout the interior of any $C_{\rho'}$, $\rho' > \rho$. If a sequence of harmonic*

polynomials $p_n(x, y)$ *of respective degrees n satisfies* (10), *then the sequence* $p_n(x, y)$ *is said to converge to* $u(x, y)$ *on C maximally, or with the greatest geometric degree of convergence.*

Whenever C and $u(x, y)$ are given, it follows from Theorem 4 that there exists a sequence $p_n(x, y)$ converging maximally to $u(x, y)$ on C. The property of maximal convergence is shared by many sequences of polynomials with extremal properties (see §12 below), so we study in some detail the implications of maximal convergence.

THEOREM 8. *Let C be a closed limited point set consisting of a finite number of mutually exterior closed Jordan regions. Let the function* $u(x, y)$ *be single-valued and harmonic throughout the interior of* C_ρ *but not throughout the interior of any* $C_{\rho'}$, $\rho' > \rho$. *Let the sequence of harmonic polynomials* $p_n(x, y)$ *of respective degrees n be given. A necessary and sufficient condition that the sequence* $p_n(x, y)$ *converge maximally to* $u(x, y)$ *on C is that the sequence* $p_n(x, y)$ *converge maximally to* $u(x, y)$ *on the closed interior of every* C_σ, $\sigma < \rho$.

Let the sequence $p_n(x, y)$ converge maximally to $u(x, y)$ on C. By the relation $[C_\sigma]_{\rho/\sigma} = C_\rho$, maximal convergence of the sequence $p_n(x, y)$ to $u(x, y)$ on the closed interior of C_σ follows from the Corollary to Theorem 5. To be sure, maximal convergence on the closed interior of C_σ is as yet strictly defined only in the case that C_σ has no multiple points. But definition and proof extend readily (compare §10 below) to the general case.

Conversely, let us suppose for every $\sigma < \rho$

$$\overline{\lim_{n\to\infty}}\, [\mu_n(\sigma)]^{1/n} = \sigma/\rho, \qquad \mu_n(\sigma) = \max\,[\,|\, u(x, y) - p_n(x, y)\,|, (x, y) \text{ on } C_\sigma].$$

We obviously have $\mu_n \leqq \mu_n(\sigma)$, whence for every $\sigma < \rho$

$$\overline{\lim_{n\to\infty}}\, \mu_n^{1/n} \leqq \sigma/\rho.$$

Here we allow σ to approach unity and notice that by Theorem 4 the left-hand member cannot be less than $1/\rho$; we obtain (10).

Regions of uniform convergence of sequences converging maximally can be determined with a fair degree of completeness:

THEOREM 9. *Let C,* $u(x, y)$, *and* $p_n(x, y)$ *satisfy the conditions of the italicized definition. Then on and within every* $C_\sigma(\sigma < \rho)$ *the sequence* $p_n(x, y)$ *converges uniformly to* $u(x, y)$; *but the sequence* $p_n(x, y)$ *converges uniformly in no region containing in its interior a point of* C_ρ.

The first part of Theorem 9 is included in Theorem 8; we proceed to study the second part. Suppose the sequence $p_n(x, y)$ converges uniformly in some simply-connected region D containing in its interior a point (x_1, y_1) of C_ρ; we shall reach a contradiction. The region D overlaps C_ρ, throughout the interior of which $p_n(x, y)$ converges to $u(x, y)$, so the limit of $p_n(x, y)$ may be denoted by $u(x, y)$ in D even exterior to C_ρ.

If we define as usual

$$f'(z) = \frac{\partial u}{\partial x} - i\frac{\partial u}{\partial y}, \qquad P'_n(z) = \frac{\partial p_n}{\partial x} - i\frac{\partial p_n}{\partial y},$$

it follows from Theorem 1 that the equation

$$\lim_{n\to\infty} P'_n(z) = f'(z)$$

holds uniformly in some neighborhood of (x_1, y_1). It then follows from Theorem 7 and from a general result on sequences of polynomials in the complex variable,[8] that this equation holds uniformly on any closed set interior to some $C_{\rho'}$, $\rho' > \rho$. It follows from Theorem 3 as in the proof of (16) that the sequence $p_n(x, y)$ converges throughout the interior of $C_{\rho'}$, uniformly on any closed set interior to $C_{\rho'}$, and hence that $u(x, y)$ can be harmonically extended from C so as to be single-valued and harmonic throughout the interior of $C_{\rho'}$, contrary to hypothesis. This contradiction completes the proof of Theorem 9.

By inspection of the proof just given we can formulate the slightly more general

COROLLARY. *Let C be a closed limited point set consisting of a finite number of mutually exterior closed Jordan regions. Let the function $u(x, y)$ be defined on C, and let the harmonic polynomials $p_n(x, y)$ of respective degrees n satisfy* (10). *If the sequence $p_n(x, y)$ converges uniformly in a region having in its interior a point of C_ρ, then for a suitable value of $\sigma < \rho$, and for some closed region D containing a point of C_ρ in its interior, we have*

$$\overline{\lim_{n\to\infty}} \{\max [| u(x, y) - p_n(x, y) |, (x, y) \text{ in } D]\}^{1/n} \leqq \sigma/\rho;$$

the sequence $p_n(x, y)$ converges throughout the interior of some $C_{\rho'}$, $\rho' > \rho$, uniformly on any closed set interior to $C_{\rho'}$. Thus $u(x, y)$ is single-valued and harmonic throughout the interior of $C_{\rho'}$.

Theorem 9 asserts the impossibility of uniform convergence of the sequence

---

[8] *Let C be an arbitrary closed limited point set whose complement is connected and regular. Let $\Phi(z)$ be defined on C and let the polynomials $\pi_n(z)$ of respective degrees n satisfy the condition*

$$\overline{\lim_{n\to\infty}} \lambda_n^{1/n} = 1/\rho < 1, \qquad \lambda_n = \max [| \Phi(z) - \pi_n(z) |, z \text{ on } C].$$

If the sequence $\pi_n(z)$ converges uniformly in some closed region containing a point of C_ρ in its interior, then for a suitable value of $\sigma < \rho$, and for some closed region D containing a point of C_ρ in its interior, we have

$$\overline{\lim_{n\to\infty}} \{\max [| \Phi(z) - \pi_n(z) |, z \text{ in } D]\}^{1/n} \leqq \sigma/\rho;$$

the sequence $\pi_n(z)$ converges throughout the interior of some $C_{\rho'}$, $\rho' > \rho$, uniformly on any closed set interior to $C_{\rho'}$.

All the material for the proof of this theorem is given [op. cit., §4.8], but a corollary of the theorem is formulated rather than the theorem itself.

$p_n(x, y)$ in any *region* containing in its interior a point of C_ρ. Nevertheless the sequence $p_n(x, y)$ may converge uniformly on a Jordan arc which is partly interior and partly exterior to C_ρ. For instance let us use polar coördinates (r, θ), and let us set

$$p_n(x, y) \equiv \sum_{m=1}^{n} \frac{1}{3^m} r^m \sin(2^m \theta), \qquad u(x, y) \equiv \lim_{n\to\infty} p_n(x, y).$$

We may choose C as the unit circle: $r \leqq 1$, so C_R is the curve $r = R$. We obviously have $\rho \geqq 3$. But in the point $r = 3$, $\theta = \pi/3$, the sequence $p_n(x, y) - p_{n-1}(x, y)$ fails to approach zero with $1/n$, the sequence $p_n(x, y)$ cannot converge, hence the point $r = 3$, $\theta = \pi/3$ cannot lie interior to C_ρ. This establishes the inequality $\rho \leqq 3$ and the equation $\rho = 3$. Nevertheless we see by inspection that the sequence $p_n(x, y)$ converges uniformly on every ray $\theta = m\pi/2^q$, $0 \leqq r < \infty$, where m and q are integers.

A result somewhat similar to the Corollary to Theorem 9 is

THEOREM 10. *Let C be a closed limited point set consisting of mutually exterior closed Jordan regions $J_1, J_2, \cdots, J_k$. Let the function $u(x, y)$ be defined on C, and let the harmonic polynomials $p_n(x, y)$ of respective degrees n satisfy* (10). *If for a single value of j we have*

$$\overline{\lim_{n\to\infty}} \{\max [|u(x, y) - p_n(x, y)|, (x, y) \text{ on } J_j]\}^{1/n} \leqq 1/\rho'' < 1/\rho, \tag{25}$$

then the sequence $p_n(x, y)$ converges throughout the interior of some $C_{\rho'}$, $\rho' > \rho$, uniformly on any closed set interior to $C_{\rho'}$. Thus $u(x, y)$ is single-valued and harmonic throughout the interior of $C_{\rho'}$.

In particular inequality (25) *cannot hold if the sequence $p_n(x, y)$ converges maximally to $u(x, y)$ on C, where ρ has the usual significance.*

By Theorem 5 we have from inequality (25) for at least one j

$$\overline{\lim_{n\to\infty}} \{\max [| u(x, y) - p_n(x, y) |, (x, y) \text{ on } [J_j]_{\rho''/\rho}]\}^{1/n} \leqq 1/\rho.$$

Let C' denote the point set composed of C plus the closed interior of $[J_j]_{\rho''/\rho}$. If ρ'' is suitably chosen, the complement of C' is connected and regular. From (10) and the inequality just proved we may write

$$\overline{\lim_{n\to\infty}} \{\max [| u(x, y) - p_n(x, y) |, (x, y) \text{ on } C']\}^{1/n} \leqq 1/\rho.$$

It follows from Theorem 4 that the sequence $p_n(x, y)$ converges to $u(x, y)$ interior to C'_ρ, uniformly on any closed set interior to C'_ρ; hence $u(x, y)$ is single-valued and harmonic throughout the interior of C'_ρ. But the locus C'_ρ contains in its interior C_ρ and hence also some $C_{\rho'}$, so the proof is complete.

The proof of Theorem 10 requires only minor changes if (25) is replaced by

$$\overline{\lim_{n\to\infty}} \{\max [| u(x, y) - p_n(x, y) |, (x, y) \text{ in } D]\}^{1/n} \leqq 1/\rho'' < 1/\rho, \tag{26}$$

where D is a closed Jordan region to which belongs either a point exterior to C or a point on the boundary of C; the conclusion of Theorem 10 remains valid.

Consequently, if the sequence $p_n(x, y)$ converges maximally to $u(x, y)$ on C, where ρ has the usual significance, inequality (26) is impossible for such a region D. Under these conditions on $u(x, y)$, $p_n(x, y)$, C, and ρ, and if D is an arbitrary closed Jordan region which is contained in C but to which belongs a boundary point of D, then the left-hand member of (26) is equal to $1/\rho$; if D is an arbitrary closed Jordan region whose interior lies interior to some $C_\sigma(\sigma < \rho)$ but one of whose boundary points lies on C_σ, then the left-hand member of (26) is equal to σ/ρ.

This more general proposition has immediate application, which we shall not mention in further detail, to many of the sequences considered in §§7–9 below.[9]

7. Differentiation of maximal sequences. It is natural to inquire whether the operations of differentiation and integration alter the property of maximal convergence. We shall now discuss these questions, always with the notation and hypothesis of the italicized definition of maximal convergence.

THEOREM 11. *Let the sequence $p_n(x, y)$ converge maximally to $u(x, y)$ on C. Then we have*

$$(27)\quad \begin{aligned} &\overline{\lim_{n\to\infty}}\, \mu_{1n}^{1/n} = 1/\rho, \quad \mu_{1n} = \max\,[\,|\,\partial u/\partial x - \partial p_n/\partial x\,|, \quad (x, y) \text{ on } C],\\ &\overline{\lim_{n\to\infty}}\, \mu_{2n}^{1/n} = 1/\rho, \quad \mu_{2n} = \max\,[\,|\,\partial u/\partial y - \partial p_n/\partial y\,|, \quad (x, y) \text{ on } C]. \end{aligned}$$

In particular if $\partial u/\partial x$ [or $\partial u/\partial y$] is single-valued and harmonic throughout the interior of no $C_{\rho'}$, $\rho' > \rho$, and therefore always in the case $k = 1$, then the sequence $\partial p_n/\partial x$ [or $\partial p_n/\partial y$] converges maximally to $\partial u/\partial x$ [or $\partial u/\partial y$] on C.

---

[9] The analogue of the more general result can be proved in an analogous manner:

Let C be a closed limited point set whose complement is connected and regular. Let $f(z)$ be defined on C, and let polynomials $p_n(z)$ in z of respective degrees n satisfy the relations

$$(26')\quad \begin{aligned} &\overline{\lim_{n\to\infty}}\, \{\max\,[\,|\,f(z) - p_n(z)\,|,\ z \text{ on } C]\}^{1/n} = 1/\rho < 1,\\ &\overline{\lim_{n\to\infty}}\, \{\max\,[\,|\,f(z) - p_n(z)\,|,\ z \text{ on } D]\}^{1/n} = 1/\rho'' < 1/\rho, \end{aligned}$$

where D is a closed set whose complement is connected and regular, and where D contains either a point exterior to C or a point on the boundary of C. Then the sequence $p_n(z)$ converges throughout the interior of some $C_{\rho'}$, $\rho' > \rho$, uniformly on any closed set interior to $C_{\rho'}$. Thus $f(z)$ is single-valued and analytic throughout the interior of $C_{\rho'}$.

Consequently, if the sequence $p_n(z)$ converges maximally to $f(z)$ on C, where ρ has the usual significance [op. cit., §4.7] *inequality (26′) is impossible for such a set D. Under these conditions on $f(z)$, $p_n(z)$, C, and ρ, and if D (closed, with complement connected and regular) belongs to C but contains a point of the boundary of C, then the left-hand member of (26′) is equal to $1/\rho$; if every point of D (closed, with complement connected and regular) lies on or interior to some C_σ ($\sigma < \rho$), but if at least one boundary point of D lies on C_σ, then the left-hand member of (26′) is equal to σ/ρ.*

From Theorem 7 we have

$$\overline{\lim_{n\to\infty}}\, \mu_{1n}^{1/n} \leqq 1/\rho, \qquad \overline{\lim_{n\to\infty}}\, \mu_{2n}^{1/n} \leqq 1/\rho. \tag{28}$$

Neither strong inequality can hold here. Suppose for instance we have

$$\overline{\lim_{n\to\infty}}\, \mu_{1n}^{1/n} = 1/\rho' < 1/\rho, \qquad \overline{\lim_{n\to\infty}}\, \mu_{2n}^{1/n} \leqq 1/\rho. \tag{29}$$

The function $\partial p_n/\partial x$ is a harmonic polynomial of degree $n - 1$, or if we prefer of degree n. It follows from Theorem 4 that we have

$$\lim_{n\to\infty} \frac{\partial p_n(x, y)}{\partial x} = \frac{\partial u(x, y)}{\partial x}$$

throughout the interior of $C_{\rho'}$, uniformly on any closed set interior to $C_{\rho'}$. Each of the finite regions into which $C_{\rho'}$ separates the plane contains in its interior at least one of the closed Jordan regions J_j composing C, and interior to each region J_j we choose a point (a_j, b_j). By the maximal convergence of $p_n(x, y)$ to $u(x, y)$ on C we have

$$\lim_{n\to\infty} p_n(a_j, b_j) = u(a_j, b_j).$$

By the second of relations (29) we have

$$\lim_{n\to\infty} \frac{\partial p_n(a_j, b_j)}{\partial y} = \frac{\partial u(a_j, b_j)}{\partial y}.$$

From Theorem 3 it follows that $p_n(x, y)$ approaches $u(x, y)$ throughout the interior of $C_{\rho'}$, uniformly on any closed set interior to $C_{\rho'}$, and hence that $u(x, y)$ is single-valued and harmonic throughout the interior of $C_{\rho'}$, $\rho' > \rho$. This contradicts the definition of ρ, and thereby proves the impossibility of (29). In a precisely similar way it can be shown that the simultaneous relations

$$\overline{\lim_{n\to\infty}}\, \mu_{1n}^{1/n} \leqq 1/\rho, \qquad \overline{\lim_{n\to\infty}}\, \mu_{2n}^{1/n} = 1/\rho' < 1/\rho$$

are impossible, so equations (27) are established.

If $\partial u/\partial x$ [or $\partial u/\partial y$] is harmonic at a point, so also is $\partial u/\partial y$ [or $\partial u/\partial x$], hence so also is $u(x, y)$. Thus in the case $k = 1$ the function $u(x, y)$ has a singularity on C_ρ, and both the functions $\partial u/\partial x$ and $\partial u/\partial y$ have singularities on C_ρ. Theorem 11 is completely proved.

It is to be noticed that under the hypothesis of Theorem 11 with $k > 1$ we cannot assert the maximal convergence on C of the sequences $\partial p_n/\partial x$ and $\partial p_n/\partial y$ to the respective functions $\partial u/\partial x$ and $\partial u/\partial y$. For instance let us choose $k = 2$, $u(x, y)$ equal to 1 and 2 in J_1 and J_2 respectively. Then we have $\partial u/\partial x \equiv 0$, $\partial u/\partial y \equiv 0$, $\rho < \infty$, so by (27) the sequences $\partial p_n/\partial x$ and $\partial p_n/\partial y$ do not converge maximally on C.

It may also occur in Theorem 11 that $\partial u/\partial x$ is single-valued and harmonic throughout the interior of no $C_{\rho'}$, $\rho' > \rho$, whereas $\partial u/\partial y$ does not possess that

property. This situation arises for instance if C_ρ separates the plane into precisely two finite regions, with $\partial u/\partial x$ zero and unity in these respective regions, and $\partial u/\partial y$ zero in both regions.

Theorem 11 establishes more specific results than Theorems 1 and 7, under a more restricted hypothesis.

Under the general conditions of Theorem 11, the sequence $\partial p_n/\partial x$ [or $\partial p_n/\partial y$] cannot converge uniformly in a region containing in its interior a point of C_ρ. For if this were to occur, the sequences $\partial p_n/\partial y$ [or $\partial p_n/\partial x$] and $p_n(x, y)$ would also converge uniformly in such a region, by Theorems 2 and 3 respectively, in contradiction to Theorem 9. Thus we have proved the

COROLLARY. *Under the hypothesis of Theorem 11, neither of the sequences $\partial p_n/\partial x$, $\partial p_n/\partial y$ can converge uniformly in a region containing in its interior a point of C_ρ.*

The method used in the proof of Theorem 11 and its Corollary applies by iteration to the study of the higher partial derivatives of $p_n(x, y)$ and $u(x, y)$. In each case we have

$$\varlimsup_{n\to\infty}\left\{\max\left[\left|\frac{\partial^{i+j}u}{\partial x^i\,\partial y^j}-\frac{\partial^{i+j}p_n}{\partial x^i\,\partial y^j}\right|,\ (x,y)\text{ on }C\right]\right\}^{1/n}=\frac{1}{\rho}. \tag{30}$$

The sequence $\partial^{i+j}p_n/\partial x^i\partial y^j$ can converge uniformly in no region containing in its interior a point of C_ρ. If $\partial^{i+j}u/\partial x^i\partial y^j$ is single-valued and harmonic throughout the interior of no $C_{\rho'}$, $\rho' > \rho$ (this always occurs if $k = 1$), equation (30) indicates maximal convergence on C of the sequence $\partial^{i+j}p_n/\partial x^i\partial y^j$ to the function $\partial^{i+j}u/\partial x^i\partial y^j$.

It is worth mentioning in connection with Theorem 11 that not merely equations (27) are valid but also the equations

$$\begin{aligned}&\varlimsup_{n\to\infty}\left\{\max\left[\left|\frac{\partial u}{\partial x}-\frac{\partial p_n}{\partial x}\right|,\ (x,y)\text{ on }J_j\right]\right\}^{1/n}=1/\rho,\\ &\varlimsup_{n\to\infty}\left\{\max\left[\left|\frac{\partial u}{\partial y}-\frac{\partial p_n}{\partial y}\right|,\ (x,y)\text{ on }J_j\right]\right\}^{1/n}=1/\rho,\end{aligned} \tag{31}$$

for every j. If equations (31) do not hold, at least one of those equations (say the former) is to be replaced by the corresponding inequality (the left-hand member is less than the right-hand member) for some value of j. From Theorem 10 follows the convergence of the sequence $\partial p_n/\partial x$ to the function $\partial u/\partial x$ throughout the interior of some $C_{\rho'}$, $\rho' > \rho$, uniformly on any closed set interior to $C_{\rho'}$. This is impossible, by the method used in the proof of Theorem 11.

Equations similar to (31) hold also for the higher partial derivatives of $u(x, y)$ and $p_n(x, y)$, and may be similarly established.[10]

---

[10] A similar proof yields the analogue:

Let C be a closed limited point set whose complement is connected and regular. Let C_1 be a closed subset of C which contains at least one point of the boundary of C and whose comple-

From Theorem 7 we have

$$\overline{\lim_{n\to\infty}}\ \mu_{1n}^{1/n} \leqq 1/\rho, \qquad \overline{\lim_{n\to\infty}}\ \mu_{2n}^{1/n} \leqq 1/\rho. \tag{28}$$

Neither strong inequality can hold here. Suppose for instance we have

$$\overline{\lim_{n\to\infty}}\ \mu_{1n}^{1/n} = 1/\rho' < 1/\rho, \qquad \overline{\lim_{n\to\infty}}\ \mu_{2n}^{1/n} \leqq 1/\rho. \tag{29}$$

The function $\partial p_n/\partial x$ is a harmonic polynomial of degree $n - 1$, or if we prefer of degree n. It follows from Theorem 4 that we have

$$\lim_{n\to\infty} \frac{\partial p_n(x, y)}{\partial x} = \frac{\partial u(x, y)}{\partial x}$$

throughout the interior of $C_{\rho'}$, uniformly on any closed set interior to $C_{\rho'}$. Each of the finite regions into which $C_{\rho'}$ separates the plane contains in its interior at least one of the closed Jordan regions J_j composing C, and interior to each region J_j we choose a point (a_j, b_j). By the maximal convergence of $p_n(x, y)$ to $u(x, y)$ on C we have

$$\lim_{n\to\infty} p_n(a_j, b_j) = u(a_j, b_j).$$

By the second of relations (29) we have

$$\lim_{n\to\infty} \frac{\partial p_n(a_j, b_j)}{\partial y} = \frac{\partial u(a_j, b_j)}{\partial y}.$$

From Theorem 3 it follows that $p_n(x, y)$ approaches $u(x, y)$ throughout the interior of $C_{\rho'}$, uniformly on any closed set interior to $C_{\rho'}$, and hence that $u(x, y)$ is single-valued and harmonic throughout the interior of $C_{\rho'}$, $\rho' > \rho$. This contradicts the definition of ρ, and thereby proves the impossibility of (29). In a precisely similar way it can be shown that the simultaneous relations

$$\overline{\lim_{n\to\infty}}\ \mu_{1n}^{1/n} \leqq 1/\rho, \qquad \overline{\lim_{n\to\infty}}\ \mu_{2n}^{1/n} = 1/\rho' < 1/\rho$$

are impossible, so equations (27) are established.

If $\partial u/\partial x$ [or $\partial u/\partial y$] is harmonic at a point, so also is $\partial u/\partial y$ [or $\partial u/\partial x$], hence so also is $u(x, y)$. Thus in the case $k = 1$ the function $u(x, y)$ has a singularity on C_ρ, and both the functions $\partial u/\partial x$ and $\partial u/\partial y$ have singularities on C_ρ. Theorem 11 is completely proved.

It is to be noticed that under the hypothesis of Theorem 11 with $k > 1$ we cannot assert the maximal convergence on C of the sequences $\partial p_n/\partial x$ and $\partial p_n/\partial y$ to the respective functions $\partial u/\partial x$ and $\partial u/\partial y$. For instance let us choose $k = 2$, $u(x, y)$ equal to 1 and 2 in J_1 and J_2 respectively. Then we have $\partial u/\partial x \equiv 0$, $\partial u/\partial y \equiv 0$, $\rho < \infty$, so by (27) the sequences $\partial p_n/\partial x$ and $\partial p_n/\partial y$ do not converge maximally on C.

It may also occur in Theorem 11 that $\partial u/\partial x$ is single-valued and harmonic throughout the interior of no $C_{\rho'}$, $\rho' > \rho$, whereas $\partial u/\partial y$ does not possess that

property. This situation arises for instance if C_ρ separates the plane into precisely two finite regions, with $\partial u/\partial x$ zero and unity in these respective regions, and $\partial u/\partial y$ zero in both regions.

Theorem 11 establishes more specific results than Theorems 1 and 7, under a more restricted hypothesis.

Under the general conditions of Theorem 11, the sequence $\partial p_n/\partial x$ [or $\partial p_n/\partial y$] cannot converge uniformly in a region containing in its interior a point of C_ρ. For if this were to occur, the sequences $\partial p_n/\partial y$ [or $\partial p_n/\partial x$] and $p_n(x, y)$ would also converge uniformly in such a region, by Theorems 2 and 3 respectively, in contradiction to Theorem 9. Thus we have proved the

COROLLARY. *Under the hypothesis of Theorem* 11, *neither of the sequences* $\partial p_n/\partial x$, $\partial p_n/\partial y$ *can converge uniformly in a region containing in its interior a point of* C_ρ.

The method used in the proof of Theorem 11 and its Corollary applies by iteration to the study of the higher partial derivatives of $p_n(x, y)$ and $u(x, y)$. In each case we have

$$\overline{\lim_{n\to\infty}}\left\{\max\left[\left|\frac{\partial^{i+j}u}{\partial x^i\,\partial y^j}-\frac{\partial^{i+j}p_n}{\partial x^i\,\partial y^j}\right|,\ (x, y) \text{ on } C\right]\right\}^{1/n}=\frac{1}{\rho}. \tag{30}$$

The sequence $\partial^{i+j}p_n/\partial x^i\partial y^j$ can converge uniformly in no region containing in its interior a point of C_ρ. If $\partial^{i+j}u/\partial x^i\partial y^j$ is single-valued and harmonic throughout the interior of no $C_{\rho'}$, $\rho' > \rho$ (this always occurs if $k = 1$), equation (30) indicates maximal convergence on C of the sequence $\partial^{i+j}p_n/\partial x^i\partial y^j$ to the function $\partial^{i+j}u/\partial x^i\partial y^j$.

It is worth mentioning in connection with Theorem 11 that not merely equations (27) are valid but also the equations

$$\begin{aligned}&\overline{\lim_{n\to\infty}}\left\{\max\left[\left|\frac{\partial u}{\partial x}-\frac{\partial p_n}{\partial x}\right|,\ (x, y) \text{ on } J_j\right]\right\}^{1/n}=1/\rho,\\ &\overline{\lim_{n\to\infty}}\left\{\max\left[\left|\frac{\partial u}{\partial y}-\frac{\partial p_n}{\partial y}\right|,\ (x, y) \text{ on } J_j\right]\right\}^{1/n}=1/\rho,\end{aligned} \tag{31}$$

for every j. If equations (31) do not hold, at least one of those equations (say the former) is to be replaced by the corresponding inequality (the left-hand member is less than the right-hand member) for some value of j. From Theorem 10 follows the convergence of the sequence $\partial p_n/\partial x$ to the function $\partial u/\partial x$ throughout the interior of some $C_{\rho'}$, $\rho' > \rho$, uniformly on any closed set interior to $C_{\rho'}$. This is impossible, by the method used in the proof of Theorem 11.

Equations similar to (31) hold also for the higher partial derivatives of $u(x, y)$ and $p_n(x, y)$, and may be similarly established.[10]

---

[10] A similar proof yields the analogue:

Let C be a closed limited point set whose complement is connected and regular. Let C_1 *be a closed subset of C which contains at least one point of the boundary of C and whose comple-*

8. **Integration of maximal sequences.** Let $p_n(x, y)$ be a harmonic polynomial of degree n, and let $q_n(x, y)$ be a conjugate function:

$$(32) \qquad q_n(x, y) \equiv q_n(a, b) + \int_{(a,b)}^{(x,y)} \left(-\frac{\partial p_n}{\partial y} dx + \frac{\partial p_n}{\partial x} dy \right).$$

Then the functions

$$(33) \qquad \begin{aligned} P_n(x, y) &\equiv P_n(a, b) + \int_{(a,b)}^{(x,y)} (p_n(x, y)\, dx - q_n(x, y)\, dy), \\ Q_n(x, y) &\equiv Q_n(a, b) + \int_{(a,b)}^{(x,y)} (q_n(x, y)\, dx + p_n(x, y)\, dy), \end{aligned}$$

are harmonic polynomials of degree $n + 1$, for the function

$$P_n(x, y) + iQ_n(x, y) \equiv P_n(a, b) + iQ_n(a, b) + \int_{a+ib}^{x+iy} (p_n + iq_n)\, dz$$

is a polynomial in z of degree $n + 1$. The following theorem suggests itself:

THEOREM 12. *Let the sequence of harmonic polynomials $p_n(x, y)$ converge maximally to the function $u(x, y)$ in the closed Jordan region C. Let (a, b) be a fixed point interior to C. Let us introduce the notation* (32) *and* (33) *together with*

$$v(x, y) \equiv v(a, b) + \int_{(a,b)}^{(x,y)} \left(-\frac{\partial u}{\partial y} dx + \frac{\partial u}{\partial x} dy \right),$$

$$(34) \qquad \begin{aligned} U(x, y) &\equiv U(a, b) + \int_{(a,b)}^{(x,y)} (u(x, y)\, dx - v(x, y)\, dy), \\ V(x, y) &\equiv V(a, b) + \int_{(a,b)}^{(x,y)} (v(x, y)\, dx + u(x, y)\, dy), \end{aligned}$$

where the integrals are to be taken over paths interior to C_ρ (usual notation). Let us assume

$$\overline{\lim_{n\to\infty}}\, | v(a, b) - q_n(a, b) |^{1/n} \leqq 1/\rho.$$

If we have

$$\overline{\lim_{n\to\infty}}\, | U(a, b) - P_n(a, b) |^{1/n} \leqq 1/\rho,$$

---

ment is connected and regular. Let the sequence of polynomials $p_n(z)$ of respective degrees n converge maximally to $f(z)$ on C. Then we have also

$$\lim_{n\to\infty} \{\max [| f^{(j)}(z) - p_n^{(j)}(z) |, z \text{ on } C_1]\}^{1/n} = 1/\rho,$$

where ρ has the usual significance. The conclusion continues to hold if C_1 is replaced by C. The sequence $p_n^{(j)}(z)$ converges uniformly in no region containing in its interior a point of C_ρ. If $f^{(j)}(z)$ is single-valued and analytic throughout the interior of no $C_{\rho'}$, $\rho' > \rho$, then the sequence $p_n^{(j)}(z)$ converges maximally to $f^{(j)}(z)$ on C.

then the sequence $P_n(x, y)$ converges maximally to $U(x, y)$ on C. If we have

$$\overline{\lim_{n\to\infty}} \mid V(a, b) - Q_n(a, b) \mid^{1/n} \leqq 1/\rho,$$

then the sequence $Q_n(x, y)$ converges maximally to $V(x, y)$ on C.

Since $u(x, y)$ is harmonic throughout the interior of C_ρ but not throughout the interior of any $C_{\rho'}$, $\rho' > \rho$, the same is true of $v(x, y)$ and of $U(x, y)$ and $V(x, y)$. Theorem 12 is then a consequence of Theorems 2, 3, and 6.[11]

Theorem 12 does not extend at once to the case of a point set C which consists of several mutually exterior closed Jordan regions $J_1, J_2, \cdots, J_k$. To be sure, under these conditions a point (a_j, b_j) can be chosen interior to each J_j, and functions $v_j(x, y)$, $q_{nj}(x, y)$, $U_j(x, y)$, $V_j(x, y)$, $P_{nj}(x, y)$, $Q_{nj}(x, y)$ can be defined by equations analogous to those already used. But these functions $P_{nj}(x, y)$ and $Q_{nj}(x, y)$ may depend effectively on j; it may be impossible to define such functions as polynomials independent of j.

We illustrate the fact just mentioned by an example already used elsewhere [op. cit., pp. 82–83]. Let C be the point set $\mid z^2 - 1 \mid \leqq 1/4$, and let $u(x, y)$ in the left-hand oval of the curve $\mid z^2 - 1 \mid = 1/2$ be identically zero and in the right-hand oval be $\frac{1}{2} \log [(x - \frac{1}{2}\sqrt{2})^2 + y^2]$. The function $u(x, y)$ is harmonic throughout the interior of C_ρ but not throughout the interior of any $C_{\rho'}$, $\rho' > \rho$, where $\rho = 2^{1/2}$. Let the polynomials $p_n(x, y)$ be found by taking the real part of the expansion of the function $f(z) \equiv u(x, y) + iv(x, y)$ (where $v(x, y)$ is conjugate to $u(x, y)$ on C) in a Jacobi series:

$$f(z) \equiv a_0 + a_1(z + 1) + a_2(z^2 - 1) + a_3(z^2 - 1)(z + 1) + \cdots, \overline{\lim_{n\to\infty}} \mid a_n \mid^{1/n} = 2^{1/2}.$$

We may choose (a_1, b_1) as $(-1, 0)$ and (a_2, b_2) as $(1, 0)$. We then have by (33)

$$P_n(1, 0) - P_n(-1, 0) = \int_{-1}^{1} p_n(x, y)\, dx,$$

and this sequence diverges [op. cit., p. 83]. That is to say, it is impossible to find a harmonic polynomial $P_n(x, y)$ represented by (33) such that the two sequences $P_n(1, 0)$ and $P_n(-1, 0)$ simultaneously converge.

It may naturally occur in particular instances when C consists of mutually

---

[11] Suppose we have $A_n \geqq 0$, $B_n \geqq 0$,

$$\overline{\lim_{n\to\infty}} A_n^{1/n} \leqq 1/\rho, \qquad \overline{\lim_{n\to\infty}} B_n^{1/n} \leqq 1/\rho, \qquad \rho > 1.$$

If $R < \rho$ is arbitrary, we then have for sufficiently large n

$$A_n \leqq 1/R^n, \qquad B_n \leqq 1/R^n, \qquad A_n + B_n \leqq 2/R^n,$$

$$\overline{\lim_{n\to\infty}} [A_n + B_n]^{1/n} \leqq 1/R,$$

whence by the arbitrary character of $R < \rho$,

$$\overline{\lim_{n\to\infty}} [A_n + B_n]^{1/n} \leqq 1/\rho.$$

exterior closed Jordan regions $J_1, J_2, \cdots, J_k$ that functions $v_j(x, y)$, $U_j(x, y)$, $V_j(x, y)$, and polynomials $q_{nj}(x, y)$, $P_{nj}(x, y)$, $Q_{nj}(x, y)$ can be defined in J_j so as to be independent of j. Under such circumstances Theorem 12 can be directly extended. In any case there can be established the

COROLLARY. *Let the sequence of harmonic polynomials $p_n(x, y)$ converge maximally to the function $u(x, y)$ on the point set C consisting of the mutually exterior closed Jordan regions $J_1, J_2, \ldots, J_k$. Let the point (a_j, b_j) lie interior to J_j. Let the functions $v_j(x, y)$, $q_{nj}(x, y)$, $U_j(x, y)$, $V_j(x, y)$, $P_{nj}(x, y)$, $Q_{nj}(x, y)$ be defined by equations analogous to those used in Theorem 12, where we have for every j*

$$\overline{\lim_{n\to\infty}} \mid v_j(a_j, b_j) - q_{nj}(a_j, b_j) \mid^{1/n} \leqq 1/\rho.$$

If for a particular j we have

$$\overline{\lim_{n\to\infty}} \mid U_j(a_j, b_j) - P_{nj}(a_j, b_j) \mid^{1/n} \leqq 1/\rho,$$

then we have for that j

$$\overline{\lim_{n\to\infty}} \{\max \mid U_j(x, y) - P_{nj}(x, y) \mid, (x, y) \text{ on } J_j\}^{1/n} = 1/\rho. \tag{35}$$

If for a particular j we have

$$\overline{\lim_{n\to\infty}} \mid V_j(a_j, b_j) - Q_{nj}(a_j, b_j) \mid^{1/n} \leqq 1/\rho,$$

then we have for that j

$$\overline{\lim_{n\to\infty}} \{\max \mid V_j(x, y) - Q_{nj}(x, y) \mid^{1/n}, (x, y) \text{ on } J_j\}^{1/n} = 1/\rho. \tag{36}$$

Neither the sequence $P_{nj}(x, y)$ nor the sequence $Q_{nj}(x, y)$ converges uniformly in a region having in its interior a point of C_ρ.

The method of proof of Theorem 12 shows that the left-hand member of (35) is less than or equal to the right-hand member. This relation cannot be *less than*, by Theorems 1, 6, and 10. Thus (35) is established, and (36) can be established similarly.

The last part of the Corollary follows from Theorem 1 and Theorem 9.

The conclusions and methods of proof of both Theorem 12 and the Corollary extend to iterated integrals.

9. Related sequences of polynomials in z. We have repeatedly indicated the analogy between maximal convergence of harmonic polynomials and maximal convergence of polynomials in z. We undertake now to study the relation in more detail, as concerns the obtaining of one kind of maximal sequence from the other.

THEOREM 13. *Let C be a closed limited point set consisting of the mutually exterior closed Jordan regions $J_1, J_2, \cdots, J_k$. Let the sequence of polynomials $p_n(z) \equiv p_n(x, y) + iq_n(x, y)$ of respective degrees n converge maximally on C to the*

function $f(z) \equiv u(x, y) + iv(x, y)$. *Let* $f(z)$ *be single-valued and analytic throughout the interior of* C_ρ *but not throughout the interior of any* $C_{\rho'}$, $\rho' > \rho$. *Then we have*

$$(37)\qquad \begin{aligned} &\overline{\lim_{n\to\infty}}\, \{\max [| u(x, y) - p_n(x, y) |, (x, y) \text{ on } C]\}^{1/n} = 1/\rho, \\ &\overline{\lim_{n\to\infty}}\, \{\max [| v(x, y) - q_n(x, y) |, (x, y) \text{ on } C]\}^{1/n} = 1/\rho. \end{aligned}$$

We may even write for every j

$$(38)\qquad \begin{aligned} &\overline{\lim_{n\to\infty}}\, \{\max [| u(x, y) - p_n(x, y) |, (x, y) \text{ on } J_j]\}^{1/n} = 1/\rho, \\ &\overline{\lim_{n\to\infty}}\, \{\max [| v(x, y) - q_n(x, y) |, (x, y) \text{ on } J_j]\}^{1/n} = 1/\rho. \end{aligned}$$

Neither of the sequences $p_n(x, y)$ *and* $q_n(x, y)$ *converges uniformly in a region containing in its interior a point of* C_ρ.

In particular if $u(x, y)$ [*or* $v(x, y)$] *is single-valued and harmonic throughout the interior of no* $C_{\rho'}$, $\rho' > \rho$, *and this situation always occurs if* $k = 1$, *the sequence* $p_n(x, y)$ [*or* $q_n(x, y)$] *converges maximally to* $u(x, y)$ [*or* $v(x, y)$] *on* C.

The sequence $p_n(z)$ converges uniformly in no region containing in its interior a point of C_ρ [op. cit., p. 83]. If $p_n(x, y)$ [*or* $q_n[x, y)$] were to converge uniformly in such a region, so also would $q_n(x, y)$ [or $p_n(x, y)$], by Theorem 2, and so would $p_n(z)$. Thus neither of the sequences $p_n(x, y)$ and $q_n(x, y)$ converges uniformly in such a region.

From the relation of maximal convergence

$$\overline{\lim_{n\to\infty}}\, \{\max [| f(z) - p_n(z) |, z \text{ on } C]\}^{1/n} = 1/\rho$$

and the obvious relations

$$| u(x, y) - p_n(x, y) | \leqq | f(z) - p_n(z) |, \quad | v(x, y) - q_n(x, y) | \leqq | f(z) - p_n(z) |,$$

it follows that the left-hand members in (37) are less than or equal to the right-hand members. From Theorem 4 it follows that neither of these relations can be *less than*. Thus (37) is established. Equations (38) are a consequence of Theorem 10.

If $f(z)$ has a singularity on C_ρ, so also have *both* $u(x, y)$ and $v(x, y)$. If either of the functions $u(x, y)$ and $v(x, y)$ has a singularity on C_ρ, so also has the other, and likewise the function $f(z)$. The last part of Theorem 13 now follows, and the proof is complete.

It may readily occur in Theorem 13 that $u(x, y)$ is single-valued and harmonic throughout the interior of no $C_{\rho'}$, $\rho' > \rho$, whereas $v(x, y)$ does not possess that property; compare the illustration given under Theorem 11. But at least one of the functions $u(x, y)$ and $v(x, y)$ must possess that property.

COROLLARY. *Under the hypothesis of Theorem 13, equation (30) is valid. Indeed we have for every m*

$$\overline{\lim_{n\to\infty}}\left\{\max\left[\left|\frac{\partial^{i+j}u}{\partial x^i\,\partial y^j}-\frac{\partial^{i+j}p_n}{\partial x^i\,\partial y^j}\right|,\ (x,y)\text{ on }J_m\right]\right\}^{1/n}=\frac{1}{\rho}.$$

The sequence $\partial^{i+j}p_n/\partial x^i\,\partial y^j$ can converge uniformly in no region containing in its interior a point of C_ρ.

The proof of the Corollary is similar to the discussion of §7, and is left to the reader.

Under the hypothesis of Theorem 13, the sequences investigated in §8 can be studied by methods already developed in detail. The conclusions are precisely those of §8; both formulation and proof of the new results are left to the reader.

THEOREM 14. *Let the sequence of harmonic polynomials $p_n(x, y)$ converge maximally to the function $u(x, y)$ in the closed Jordan region C. Let (a, b) be a fixed point interior to C. Let $v(x, y)$ and $q_n(x, y)$ be conjugate to $u(x, y)$ and $p_n(x, y)$ respectively in C, where*

$$\overline{\lim_{n\to\infty}}\,|\,v(a,b)-q_n(a,b)\,|^{1/n}\leqq 1/\rho.$$

Then the sequences $q_n(x, y)$ and $p_n(z) \equiv p_n(x, y) + iq_n(x, y)$ converge maximally to $v(x, y)$ and $f(z) \equiv u(x, y) + iv(x, y)$ respectively in C.

Since $u(x, y)$ has a singularity on C_ρ, so also have $v(x, y)$ and $f(z)$. The conclusion of Theorem 14 follows from Theorem 2, and from Theorem 6 applied to the sequence $q_n(x, y)$.

Theorem 14 does not extend without further qualifications to the case of a point set C which consists of several mutually exterior closed Jordan regions, as we propose to show by the special example already used in connection with Theorem 12. In the notation of §8, we set $v(-1, 0) = v(1, 0) = 0$, $q_n(-1, 0) = q_n(1, 0) = 0$; this choice is allowable, for the Jacobi series is found by interpolation in the points $(-1, 0)$ and $(+1, 0)$ to the function $u(x, y) + iv(x, y)$. Let us denote the closed interior of the left-hand oval of the lemniscate $|z^2 - 1| = 1/4$ by J_1 and the closed interior of the right-hand oval by J_2. Let us set $(a_1, b_1) = (-1, 0)$, $(a_2, b_2) = (1, 0)$, $V_j(a_j, b_j) = Q_{nj}(a_j, b_j) = 0$. The function $f(z)$ is real on the axis of reals, so the coefficients in the Jacobi development of $f(z)$ are real. Thus we have $q_n(x, 0) \equiv 0$,

$$\int_{(-1,0)}^{(1,0)}(q_n(x,y)\,dx+p_n(x,y)\,dy)=0,$$

since the integral is independent of the path. Then *the polynomial $Q_{nj}(x, y)$ is independent of j.* Similarly we have

$$\int_{(-1,0)}^{(1,0)}(v(x,y)\,dx+u(x,y)\,dy)=0,$$

where the integral is taken along the axis of reals, and *the function $V_j(x, y)$ is independent of j.* It follows from the Corollary to Theorem 12 that the

sequence $Q_{nj}(x, y) \equiv Q_n(x, y)$ converges maximally to $V_j(x, y) \equiv V(x, y)$. On the other hand, it is impossible [§8] to determine polynomials $-P_n(x, y)$ conjugate to the polynomials $Q_n(x, y)$ so that we have simultaneous convergence even of the two sequences $-P_n(-1, 0)$ and $-P_n(1, 0)$.

An obvious extension of Theorem 14 exists if C consists now of mutually exterior closed Jordan regions $J_1, J_2, \cdots, J_k$, provided the functions $q_{nj}(x, y)$ and $v_j(x, y)$ conjugate to $p_n(x, y)$ and $u(x, y)$ in J_j are allowed to depend on j; compare the Corollary to Theorem 12. The situation is somewhat simpler with derivatives:

THEOREM 15. *Let C be a closed limited point set consisting of the mutually exterior closed Jordan regions $J_1, J_2, \cdots, J_k$. Let the sequence of harmonic polynomials $p_n(x, y)$ of respective degrees n converge maximally to $u(x, y)$ on C, let the functions $q_n(x, y)$ and $v(x, y)$ be conjugate to $p_n(x, y)$ and $u(x, y)$ respectively on C, and let us set $p_n(z) \equiv p_n(x, y) + iq_n(x, y)$, $f(z) \equiv u(x, y) + iv(x, y)$. Then we have for every j, and for $m = 1, 2, \cdots$,*

$$\overline{\lim_{n\to\infty}} \{\max [|f^{(m)}(z) - p_n^{(m)}(z)|, z \text{ on } C]\}^{1/n} = 1/\rho,$$

$$\overline{\lim_{n\to\infty}} \{\max [|f^{(m)}(z) - p_n^{(m)}(z)|, z \text{ on } J_j]\}^{1/n} = 1/\rho.$$

The sequence $p_n^{(m)}(z)$ $(n = 0, 1, 2, \cdots)$ converges uniformly in no region containing in its interior a point of C_ρ. In particular if $f^{(m)}(z)$ is single-valued and analytic throughout the interior of no $C_{\rho'}$, $\rho' > \rho$ (this situation always occurs if $k = 1$), then the sequence $p_n^{(m)}(z)$ converges maximally to $f^{(m)}(z)$ on C.

Theorem 15 follows at once from the methods of §7.

THEOREM 16. *Under the hypothesis of Theorem 12 let us set*

$$P_n(z) \equiv P_n(x, y) + iQ_n(x, y), \qquad F(z) \equiv U(x, y) + iV(x, y).$$

If we have

$$\overline{\lim_{n\to\infty}} |F(a + ib) - P_n(a + ib)|^{1/n} \leq 1/\rho,$$

then the sequence $P_n(z)$ converges maximally to $F(z)$ on C.

Theorem 16 follows from Theorem 12, and a similar result can be formulated from the Corollary to Theorem 12 for the case that C consists of a finite number of mutually exterior closed Jordan regions. Both Theorem 16 and this similar result extend to higher repeated integrals of the function $f(z) \equiv u(x, y) + iv(x, y)$.

10. **More general point sets C.** Maximal convergence for sequences of polynomials in z is defined [op. cit., p. 80] for an arbitrary point set C whose complement K is connected, and is regular in the sense that Green's function exists for K with pole at infinity. We have hitherto (§§6–9) considered as point sets C on which maximal convergence of harmonic polynomials is defined, a more restricted class, namely closed limited sets each consisting of a finite number

of mutually exterior closed Jordan regions. This restriction in the study of harmonic polynomials is partly from necessity, partly for the sake of convenience.

For instance it is not desirable to define maximal convergence of a sequence of harmonic polynomials $p_n(x, y)$ of respective degrees n on a segment C of the axis of reals. Such a sequence may converge uniformly on C to a function $u(x, y)$ in such a way that we have

$$\lim_{n \to \infty} \{\max [| u(x, y) - p_n(x, y) |, (x, y) \text{ on } C]\}^{1/n} = 0,$$

and yet converge at no point not on the axis of reals. This phenomenon occurs if we set for instance $u(x, y) \equiv 0$ on C, $p_0(x, y) \equiv 0$, $p_n(x, y) \equiv n!y$ for $n \geqq 1$. A complete set of harmonic polynomials $\{r^n \cos n\theta, r^n \sin n\theta\}$, where polar coordinates are used, is not linearly independent on C. A second disadvantage of the use of a segment of the axis of reals as C is that a harmonic function whose values are given on C is not uniquely defined by those values. The addition to such a function of a polynomial $\sum_1^N a_n r^n \sin n\theta$ does not change the values on C of the function or its harmonic character in any region of the plane.

However, it is by no means essential in the theory thus far developed to require that C should be composed of only *Jordan* regions. Let J be an arbitrary closed limited simply connected region, which therefore separates the plane into a number of regions. Denote by K that one of those regions which is infinite, let $G(x, y)$ be Green's function for K with pole at infinity, and let J_R denote generically the locus $G(x, y) = \log R > 0$ in K. If $R' > 1$ is given, there exists [op. cit., p. 28, Theorem 2; the result is essentially due to Carathéodory] a closed Jordan region J' such that J lies interior to $J'_{R'}$. It follows [op. cit., p. 81, Corollary] that the inequality for every J'

$$| f(z) - p_n(z) | \leqq M/R^n, \qquad z \text{ on } J',$$

where $p_n(z)$ is a polynomial of degree n, where $R_1 < R$ is given, and where M may depend on J', implies the inequality

$$| f(z) - p_n(z) | \leqq M_1/R_1^n, \qquad z \text{ on } J'_{R/R_1} \text{ or } z \text{ on } \bar{J},$$

where M_1 is suitably chosen, and where $\bar{J}$ is the complement of K. We mention explicitly that $\bar{J}$ may contain points exterior to J.

More generally, let C be a closed limited point set whose complement K is connected and regular, let simply connected regions $J_1, J_2, \cdots, J_k$ belong to C; let C have the property that corresponding to an arbitrary R' there exist Jordan regions $J'_1, J'_2, \cdots, J'_k$ interior to $J_1, J_2, \cdots, J_k$ respectively such that every point of C lies interior to at least one of the curves $[J'_1]_{R'}, [J'_2]_{R'}, \cdots, [J'_k]_{R'}$. Let the concept of maximal convergence (§6) be extended to include such a set C. Then *the entire discussion of §§4–9 so far as concerns a set C consisting of k mutually exterior closed Jordan regions applies with only obvious modification in the present more general case. The entire discussion of §§4–9 so far as concerns a single Jordan region applies with only obvious modifications to an arbitrary closed limited simply-connected region J.* But in such a situation as the last part of Theorem 4,

we need now to assume $u(x, y)$ harmonic not merely on J but also throughout the complement of K. The complement of K may contain points exterior to J, points which are interior to every J_R.

Hitherto we have required that C should consist of a finite number of closed Jordan regions, instead of admitting more general regions, merely for the sake of simplicity. We shall continue to make that requirement for the same reason. But the succeeding material except in obvious cases extends of itself to include the more general point sets defined above. Even that extension does not completely exhaust the range of the methods used [compare op. cit., §§5.6 and 5.7].

11. Various measures of approximation. Measures of approximation other than that used in (10) are of interest, and are now to be studied. As a matter of convenience we prove first several lemmas.

Lemma IV. *Let S be a finite region bounded by a Jordan curve Γ possessing continuous curvature, and let S_1 be an arbitrary closed subregion of S. There exists a number N' depending only on S and S_1, such that if $U(x, y)$ is harmonic interior to S and continuous in the corresponding closed region, and if we set*

$$\int_\Gamma |U(x, y)|^p\, ds = \eta^p, \qquad p > 1, \tag{39}$$

then we have

$$|U(x, y)| \leqq N'\eta, \qquad (x, y) \text{ in } S_1. \tag{40}$$

The function $U(x, y)$ is represented interior to S by Green's formula

$$U(x, y) = \frac{1}{2\pi}\int_\Gamma U(\alpha, \beta)\, \frac{\partial G}{\partial \nu}\, ds_{(\alpha,\beta)}, \tag{41}$$

where $G(x, y; \alpha, \beta)$ is Green's function for S with pole in the point (x, y) and running coördinates (α, β). The normal derivative $\partial G/\partial\nu$ is continuous on Γ. We apply Hölder's inequality

$$\left|\int F^\alpha H^{1-\alpha}\right| \leqq \left(\int |F|\right)^\alpha \left(\int |H|\right)^{1-\alpha}, \qquad 0 < \alpha < 1, \tag{42}$$

to the integral in the right-hand member of (41), with $F^\alpha \equiv U$, $H^{1-\alpha} \equiv \partial G/\partial\nu$, $\alpha = 1/p$. Inequality (40) follows at once from (39) and (41).

Lemma V. *Let S be an arbitrary limited simply connected region, and let S_1 be an arbitrary closed subregion of S. There exists a constant N'' depending only on S and S_1 such that if $U(x, y)$ is harmonic interior to S and if we set*

$$\iint_S |U(x, y)|^p\, dS = \eta^p, \qquad p > 1, \tag{43}$$

then we have

$$|U(x, y)| \leqq N''\eta, \qquad (x, y) \text{ in } S_1. \tag{44}$$

Let Q denote an arbitrary circle interior to S, and let r and (x_0, y_0) be respectively the radius and center of Q. Gauss's mean value theorem yields by an integration

$$U(x_0, y_0) = \frac{1}{\pi r^2} \int\int_Q U(x, y)\, dS.$$

Inequality (42) yields

$$| U(x_0, y_0) | \leqq \frac{1}{(\pi r^2)^{1/p}} \left(\int\int_Q | U(x, y) |^p dS \right)^{1/p}.$$

This right-hand member is by (43) not greater than $\eta/(\pi r^2)^{1/p}$. If $\delta > 0$ is less than the minimum distance from S_1 to the boundary of S, inequality (44) is valid with $N'' = 1/(\pi\delta^2)^{1/p}$.

Lemma VI. *Let S be a closed limited simply connected region, and let S_1 be an arbitrary closed subregion of S. Then there exists a constant N''' depending only on S and S_1 such that if $U(x, y)$ is harmonic in S, if S is mapped conformally onto the interior of the unit circle $\gamma : | w | = 1$, and if we set*

$$(45) \qquad \int_\gamma | U(x, y) |^p | dw | = \eta^p, \qquad p > 1,$$

then we have

$$(46) \qquad | U(x, y) | \leqq N'''\eta, \qquad (x, y) \text{ in } S_1 .$$

Let the functions $w = \Phi(z)$, $z = \Psi(w)$ (assumed fixed) map the interior of S onto the interior of γ. Then $\lim_{r\to 1,\, r<1} \Psi(re^{i\theta})$ exists for almost all values of θ; a value $\theta = \theta'$ for which this limit exists corresponds to an accessible boundary point of S, and the radius $\theta = \theta'$ of γ corresponds to a Jordan arc abutting on the boundary of S in $\Psi(e^{i\theta'})$. The limit of $U(x, y)$ along this arc is the value in the end-point. Thus the integrand in (45) is defined almost everywhere on γ; the integral exists, for the integrand on γ is the limit as r approaches unity of the continuous function defined on the circle $| w | = r < 1$ as the transform of $U(x, y)$. The integral in (45) (and in similar situations) is intended to be taken in the sense just suggested, namely that limiting values of $U(x, y)$ are to be used on γ.

Poisson's integral for $U(x, y)$ is valid for the interior of the circle $| w | = r < 1$, hence for the interior of γ. Inequality (46) follows as in the proof of Lemma IV.

Lemma IV–VI yield results on degree of convergence analogous to Theorem 4:

Theorem 17. *Let C be a closed limited point set consisting of the closed interiors of mutually exterior Jordan curves $\Gamma_1, \Gamma_2, \cdots, \Gamma_k$, each possessing continuous curvature. Let the continuous function $u(x, y)$ be defined merely on the curves Γ_j. A necessary and sufficient condition that the harmonic polynomials $p_n(x, y)$ of respective degrees n satisfy the relation*

$$(47) \qquad \overline{\lim_{n\to\infty}} \left\{ \sum_{j=1}^{k} \int_{\Gamma_j} | u(x, y) - p_n(x, y) |^p ds \right\}^{1/np} = \frac{1}{\tau} < 1, \qquad p > 1,$$

is that we have

$$\varlimsup_{n\to\infty} \{\max | u(x, y) - p_n(x, y) |, (x, y) \text{ on } C\}^{1/n} = \frac{1}{\tau}. \tag{48}$$

In the first part of Theorem 4 it is a matter of indifference whether the given function $u(x, y)$ is defined at all points of C, or merely on the boundary of C, for the given sequence $p_n(x, y)$ obviously converges uniformly on the boundary of C, hence uniformly on C; the function $u(x, y)$ is obviously harmonic in the interior points of C, continuous on C. In Theorem 17 we assume $u(x, y)$ defined merely on the curves Γ_j, so we are not yet in a position to apply Lemma IV.

If (47) is given, we may write

$$\int_{\Gamma_j} | u(x, y) - p_n(x, y) |^p ds \leqq \frac{M}{\tau_1^{np}}, \qquad 1 < \tau_1 < \tau, \tag{49}$$

where M is suitably chosen. The general inequality

$$| \chi_1 + \chi_2 |^p \leqq 2^{p-1} [| \chi_1 |^p + | \chi_2 |^p], \qquad p > 1, \tag{50}$$

implies

$$\int_{\Gamma_j} | p_n(x, y) - p_{n-1}(x, y) |^p ds \leqq \frac{M_1}{\tau_1^{np}},$$

whence by Lemma IV

$$| p_n(x, y) - p_{n-1}(x, y) | \leqq \frac{M_2}{\tau_1^n}, \qquad (x, y) \text{ on } S_j, \tag{51}$$

where S_j is an arbitrary closed set interior to Γ_j. From the first part of Theorem 4 it now follows that the sequence $p_n(x, y)$ converges throughout the interior of $[S_j]_{\tau_1}$, uniformly on any closed set interior to $[S_j]_{\tau_1}$, to some function $U(x, y)$. Hence $U(x, y)$ is defined and harmonic at all points interior to $[S_j]_{\tau_1}$, in particular [by the arbitrariness of S_j interior to Γ_j; see op. cit., p. 28] at all points on or within Γ_j. Moreover we have by Theorem 5

$$| U(x, y) - p_n(x, y) | \leqq \frac{M_3}{\tau_2^n}, \qquad (x, y) \text{ on } \Gamma_j, \tag{52}$$

where $\tau_2 < \tau_1$ is arbitrary, and where M_3 (depending on τ_2) is suitably chosen. Inequality (52) yields

$$\int_{\Gamma_j} | U(x, y) - p_n(x, y) |^p ds \leqq \frac{M_4}{\tau_2^{np}},$$

which together with (49) implies by virtue of (50)

$$\int_{\Gamma_j} | U(x, y) - u(x, y) |^p ds \leqq \frac{M_5}{\tau_2^{np}}.$$

The functions $U(x, y)$ and $u(x, y)$, being both continuous on Γ_j, are therefore identical, and it is allowable to denote by $u(x, y)$ the limit of $p_n(x, y)$ on and

within Γ_j. It is essential in the first part of Theorem 17 to require that $u(x, y)$ be continuous on Γ_j if we are to conclude the identity on Γ_j of $u(x, y)$ and the limit of the sequence $p_n(x, y)$; otherwise we can conclude only the equality of $U(x, y)$ and $u(x, y)$ almost everywhere on Γ_j. This situation is to be expected, for (47) involves the values of $u(x, y)$ only almost everywhere on Γ_j.

Inequality (52) implies the inequality

$$\overline{\lim_{n\to\infty}} \{\max [| u(x, y) - p_n(x, y) |, (x, y) \text{ on } C]\}^{1/n} \leqq 1/\tau$$

so the left-hand member of (48) is less than or equal to the right-hand member. But the relation *less than* would imply the relation *less than* instead of *equality* in (47). Thus equation (48) is established.

The converse proposition follows at once, namely that (48) implies (47). Equation (48) implies that the first member of (47) is less than or equal to the second member. But the relation *less than* would by the part of Theorem 17 already proved imply the relation *less than* instead of *equality* in (48), contrary to hypothesis.

An immediate consequence of Theorem 17 is the

COROLLARY. *Let C satisfy the conditions of Theorem 17. Let the function $u(x, y)$ be single-valued and harmonic throughout the interior of C_ρ, but not throughout the interior of any $C_{\rho'}$, $\rho' > \rho$. A necessary and sufficient condition that the sequence of harmonic polynomials $p_n(x, y)$ of respective degrees n converge maximally to $u(x, y)$ on C is*

$$\overline{\lim_{n\to\infty}} \left\{\sum_{j=1}^{k} \int_{\Gamma_j} | u(x, y) - p_n(x, y) |^p ds\right\}^{1/np} = \frac{1}{\rho}. \tag{53}$$

There exists no sequence $p_n(x, y)$ such that the left-hand member of (53) *is less than* $1/\rho$.

The proofs of the following two theorems and corollaries are so similar to the proofs just given that they are left to the reader.

THEOREM 18. *Let C be a closed limited point set consisting of the mutually exterior closed Jordan regions $J_1, J_2, \cdots, J_k$. Let the continuous function $u(x, y)$ be defined merely on C. A necessary and sufficient condition that the harmonic polynomials $p_n(x, y)$ of respective degrees n satisfy the relation*

$$\overline{\lim_{n\to\infty}} \left\{\int\int_C | u(x, y) - p_n(x, y) |^p dS\right\}^{1/np} = \frac{1}{\tau} < 1, \qquad p > 1,$$

is equation (48).

COROLLARY. *Let C satisfy the conditions of Theorem 18. Let the function $u(x, y)$ be single-valued and harmonic throughout the interior of C_ρ, but not throughout the interior of any $C_{\rho'}$, $\rho' > \rho$. A necessary and sufficient condition that the*

sequence of harmonic polynomials $p_n(x, y)$ of respective degrees n converge maximally to $u(x, y)$ on C is

$$\overline{\lim_{n\to\infty}} \left\{ \int\int_C |u(x, y) - p_n(x, y)|^p dS \right\}^{1/np} = \frac{1}{\rho}. \tag{54}$$

There exists no sequence $p_n(x, y)$ such that the left-hand member of (54) *is less than $1/\rho$.*

THEOREM 19. *Let C be a closed limited point set consisting of the mutually exterior closed Jordan regions $J_1, J_2, \cdots, J_k$. Let the region J_j be mapped onto the interior of $\gamma_j : |w_j| = 1$ by a fixed transformation. Let the continuous function $u(x, y)$ be defined* a) *on the boundary of C or* b) *on the closed set C. A necessary and sufficient condition that the harmonic polynomials $p_n(x, y)$ of respective degrees n satisfy the relation*

a) $$\overline{\lim_{n\to\infty}} \left\{ \sum_{j=1}^{k} \int_{\gamma_j} |u(x, y) - p_n(x, y)|^p |dw_j| \right\}^{1/np} = \frac{1}{\tau} < 1, \qquad p > 1,$$

or b) $$\overline{\lim_{n\to\infty}} \left\{ \sum_{j=1}^{k} \int\int_{|w_j|\leq 1} |u(x, y) - p_n(x, y)|^p dS \right\}^{1/np} = \frac{1}{\tau} < 1, \qquad p > 1,$$

is that equation (48) *be fulfilled.*

COROLLARY. *Let C satisfy the conditions of Theorem* 19, *and let the function $u(x, y)$ be single-valued and harmonic throughout the interior of C_ρ but not throughout the interior of any $C_{\rho'}$, $\rho' > \rho$. Equivalent necessary and sufficient conditions that the sequence of harmonic polynomials $p_n(x, y)$ of respective degrees n converge maximally to $u(x, y)$ on C are*

$$\begin{aligned} &\overline{\lim_{n\to\infty}} \left\{ \sum_{j=1}^{k} \int_{\gamma_j} |u(x, y) - p_n(x, y)|^p |dw_j| \right\}^{1/np} = \frac{1}{\rho}, \qquad p > 1, \\ &\overline{\lim_{n\to\infty}} \left\{ \sum_{j=1}^{k} \int\int_{|w_j|\leq 1} |u(x, y) - p_n(x, y)|^p dS \right\}^{1/np} = \frac{1}{\rho}, \qquad p > 1. \end{aligned} \tag{55}$$

There exists no sequence $p_n(x, y)$ such that the left-hand member of either equation in (55) *is less than $1/\rho$.*

12. **Best approximation.** Let the point set C consist of mutually exterior closed Jordan regions $J_1, J_2, \cdots, J_k$, and let the function $u(x, y)$ be harmonic on C. The *approximation* to $u(x, y)$ on C of a harmonic polynomial $p_n(x, y)$ can be measured in any one of a number of ways, such as:

i) $$\max\, [n(x, y)\, |u(x, y) - p_n(x, y)|,\ (x, y) \text{ on } C];$$

ii) $$\sum_{j=1}^{k} \int_{\Gamma_j} n(x, y)\, |u(x, y) - p_n(x, y)|^p ds, \qquad p > 1,$$

where each region J_j is assumed to be bounded by a curve Γ_j which has continuous curvature;

iii) $$\int\int_C n(x, y) \,|\, u(x, y) - p_n(x, y) \,|^p \, dS, \qquad p > 1,$$

iv) $$\sum_{j=1}^{k} \int_{\gamma_j} n_j(w_j) \,|\, u(x, y) - p_n(x, y) \,|^p \, |\, dw_j \,|, \qquad p > 1,$$

where the interior of J_j is mapped onto the interior of $\gamma_j : |\, w_j \,| = 1$;

v) $$\sum_{j=1}^{k} \int\int_{|w_j| \leq 1} n_j(w_j) \,|\, u(x, y) - p_n(x, y) \,|^p \, dS, \qquad p > 1,$$

where the interior of J_j is mapped onto the interior of $\gamma_j : |\, w_j \,| = 1$.

The norm function $n(x, y)$ or $n_j(w_j)$ is for the present to be considered positive and continuous where defined. When n is given, there always exists at least one harmonic polynomial of degree n *of best approximation,* that is to say, a polynomial of degree n for which the measure of approximation is at least as small as for any other polynomial of degree n. In case i), the polynomial of best approximation may or may not be unique; in cases ii)–v) the polynomial of best approximation is always unique. The latter remark follows (this is the usual proof) from a detailed use of inequality (50): if two distinct polynomials of degree n are given, which have the same approximation to $u(x, y)$ on C, half their sum is actually a polynomial of degree n of *better* approximation, and differs from the two given polynomials.

The measures of approximation iv) and v) naturally depend on the particular conformal map, not uniquely determined, of the interior of J_j onto the interior of γ_j. But each of these measures of approximation with given map and given norm function $n_j(w_j)$, is equivalent to the corresponding measure of approximation with an arbitrary map and a suitable norm function.

In the following theorem, *any* polynomial of best approximation will suffice if such a polynomial is not unique:

THEOREM 20. *Let C be a closed limited point set consisting of the mutually exterior closed Jordan regions $J_1, J_2, \cdots, J_k$. Let $u(x, y)$ be harmonic on C, and let $\Pi_n(x, y)$ be the harmonic polynomial of degree n of best approximation to $u(x, y)$ on C, in any of the senses* i)–v). *Then the sequence $\Pi_n(x, y)$ converges maximally to $u(x, y)$ on C.*

Case iii) is typical; the reader will easily make the appropriate modification for the other cases. Suppose we have

$$0 < N_1 < n(x, y) < N_2, \qquad (x, y) \text{ on } C, \tag{56}$$

where N_1 and N_2 are constants. If $p_n(x, y)$ denotes a sequence of harmonic polynomials converging maximally to $u(x, y)$ on C, we have from (54) and from the last of inequalities (56),

$$\overline{\lim_{n\to\infty}} \left\{ \int\int_C n(x, y) \,|\, u(x, y) - p_n(x, y) \,|^p \, dS \right\}^{1/np} \leq \frac{1}{\rho}.$$

For the polynomials $\Pi_n(x, y)$ of *best* approximation we have a fortiori

$$\overline{\lim_{n\to\infty}} \left\{ \int\int_C n(x, y) \mid u(x, y) - \Pi_n(x, y) \mid^p dS \right\}^{1/np} \leqq \frac{1}{\rho}.$$

From the second of inequalities (56) we may write

$$\overline{\lim_{n\to\infty}} \left\{ \int\int_C \mid u(x, y) - \Pi_n(x, y) \mid^p dS \right\}^{1/np} \leqq \frac{1}{\rho}.$$

From the Corollary to Theorem 18 it follows that the inequality here is impossible, and that the sequence $\Pi_n(x, y)$ converges maximally to $u(x, y)$ on C.

For $k = 1$, $n(x, y) \equiv 1$ or $n_j(w_j) \equiv 1$, C the closed interior of a circle, method ii) reduces essentially to trigonometric approximation on the circumference, a topic which has been recently studied by Lebesgue, de la Vallée Poussin, S. Bernstein, Jackson, and others.

The measure of approximation i) is that of Tchebycheff, studied by Walsh and Julia in the case $k = 1$, $n(x, y) \equiv 1$; and by Faber in the case $k = 1$, $n(x, y) \equiv 1$, for approximation by polynomials in z. The measure ii) has been used in the case $k = 1$, $p = 2$, $n(x, y) \equiv 1$ by S. Bernstein, Brioullin, Picone, and Merriman,[12] and has been used by Szegö and Lavrentieff in the analogous situation for polynomials in the complex variable. The measure iii) has been used in the case $k = 1$, $n(x, y) \equiv 1$ by Zaremba[13] and Bergmann, and has been used by Carleman and Bochner in the analogous field of the complex variable. The measure iv) was used by Julia for harmonic polynomials in the case $k = 1$, $n_j(w_j) \equiv 1$. The most essential part of Theorem 20 for the case $k = 1$, but without the concept and detailed properties of maximal convergence, was proved in two papers by the present writer;[14] further detailed references to the literature are given there. Theorem 20 in its present form is somewhat similar to, but more specific than, results obtained by Walsh and Russell[15] for approximation by polynomials in the complex variable on a set C not necessarily connected.

In case i) with $n(x, y) \equiv 1$, it is immaterial whether approximation is measured on C or on the boundary of C. Consequently we have here a theoretical method for the solution of the Dirichlet problem. The polynomials $\Pi_n(x, y)$ can be determined merely from the values of the function $u(x, y)$ on the boundary of C. The sequence $\Pi_n(x, y)$ converges uniformly on the closed set C and thus defines $u(x, y)$ interior to C.[16] A similar remark applies to both ii) and iv).

---

[12] American Journal of Mathematics, vol. 53 (1931), pp. 589–596.

[13] Bulletin de l'Académie de Cracovie, 1909.

[14] Trans. Amer. Math. Soc., vol. 32 (1930), pp. 794–816; vol. 33 (1931), pp. 370–388.

[15] Trans. Amer. Math. Soc., vol. 36 (1934), pp. 13–28.

[16] In cases i), ii), and iv) we have here a theoretical method for the solution of the Dirichlet problem even if $u(x, y)$ is merely continuous on C, harmonic interior to C; the polynomials $\Pi_n(x, y)$ of best approximation can be determined merely from the values of the function $u(x, y)$ on the boundary of C; the sequence $\Pi_n(x, y)$ converges interior to C, uniformly on any closed set interior to C (uniformly on C itself in case i)), and thus defines

The additional relative advantage of the measure of approximation ii) lies in its naturalness and simplicity, and in the fact that for $p = 2$ the formal expansion is one in orthogonal functions. The disadvantage of it (in comparison with i), iii), iv), and v) lies in the heavy restrictions on the regions to which it applies. The advantage of iv) is that it applies to arbitrary simply connected regions, and involves the values of $u(x, y)$ only on the boundary of those regions (or on the circumference γ_j).

13. **Extensions.** We have already indicated in §10 that our results apply to a point set C much more general than that composed of a finite number of Jordan regions. The remarks of §10 apply at once to the results of §§11 and 12. But in the measure of approximation ii), it is to be noted that the boundary of C must be at least rectifiable for the integral to have a meaning; in addition, the continuity or some other requirement concerning $\partial G/\partial \nu$ on the boundary of C is essential for the reasoning as given.

In Theorem 20 we have assumed for convenience that the norm functions $n(x, y)$ and $n_j(w_j)$ are positive and continuous. It is sufficient, however, in case i) if the norm function is bounded and bounded from zero. It is sufficient in cases ii)–v) [method of op. cit., §5.7] if the norm function is integrable, together with its $(-\beta)^{\text{th}}$ power, where $\beta \geqq 1/(p-1)$.

In §§11 and 12 we have studied approximation only in the case $p > 1$. The case $p = 1$ can also be treated by similar methods and yields corresponding results. But the case $0 < p < 1$ is quite different. We are not in a position to show the failure of Theorem 20 in the latter case, but shall now show by an example the failure of Lemma IV. Let us define the continuous function $U_n(r, \theta)$ on the unit circle $\Gamma : r = 1$ by the requirements

$$U_n(1, -\pi) = 0, \qquad U_n(1, -\pi/n) = 0, \qquad U_n(1, 0) = 2n,$$
$$U_n(1, \pi/n) = 0, \qquad U_n(1, \pi) = 0,$$

where $U_n(1, \theta)$ is a linear function of θ between the successive values given. There exists a function $U_n(r, \theta)$ harmonic interior to Γ, continuous in the corresponding closed region S, which coincides with $U_n(1, \theta)$ on Γ. We have by Gauss's mean value theorem

$$U_n(0, 0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} U_n(1, \theta)\, d\theta = 1.$$

$u(x, y)$ interior to C. This remark follows from the possibility of uniform approximation of $u(x, y)$ on C by harmonic polynomials; that is to say, there exists a sequence of harmonic polynomials $p_n(x, y)$ of respective degrees n such that the measure of approximation i) approaches zero as n becomes infinite; measures i), ii), and iv) consequently approach zero for the polynomials $\Pi_n(x, y)$; we apply Lemma IV and an obvious extension of Lemma VI. The remark applies in cases i) and iv) even when the regions J_j composing C are not Jordan regions, but if the boundary of C is not the boundary of an infinite region, the function $u(x, y)$ defined continuously on the boundary of C is then not entirely arbitrary; compare Walsh, Crelle's Journal, vol. 159 (1928), pp. 197–209.

We also have $(0 < p < 1)$

$$\int_{-\pi}^{\pi} | U_n(1, \theta) |^p \, d\theta = \int_{-\pi}^{\pi} [U_n(1, \theta)]^p \, d\theta = \frac{2^{p+1}\pi}{(p + 1)n^{1-p}}.$$

This last expression approaches zero as n becomes infinite, so even for a fixed S_1 containing the origin there exists no N' such that (40) is fulfilled.

Let us consider the situation of Theorem 20, case ii). The harmonic polynomials

$$(57) \qquad 1, r \cos \theta, r \sin \theta, r^2 \cos 2\theta, r^2 \sin 2\theta, \cdots, r^N \sin N\theta$$

are linearly independent on the boundary Γ of C. For, any linear combination L of these polynomials which vanishes on Γ, vanishes throughout the interior of C, and therefore vanishes identically in the entire plane. In particular L vanishes identically on the unit circle γ, hence (since the functions 1, $\cos \theta$, $\sin \theta$, $\cdots$, $\sin N\theta$ are linearly independent on γ) all the coefficients in L are zero. It is consequently possible to orthogonalize and normalize the functions (57) on Γ with respect to the positive continuous norm function $n(x, y)$. We denote by

$$P_0(x, y), \; P_1(x, y), \; P_2(x, y), \; \cdots$$

the polynomials resulting from this process. The harmonic polynomials $P_{2n-1}(x, y)$ and $P_{2n}(x, y)$ are both of degree n.

If $u(x, y)$ is given harmonic on and within C, the harmonic polynomial $\Pi_n(x, y)$ of degree n of best approximation to $u(x, y)$ on Γ in the sense of least squares $(p = 2)$ is the sum of the first $2n + 1$ terms of the series

$$(58) \quad u(x, y) \sim \sum_{m=0}^{\infty} a_m P_m(x, y), \qquad a_m = \sum_{j=1}^{k} \int_{\Gamma_j} n(x, y) u(x, y) P_m(x, y) \, ds.$$

The actual formulas for the functions $P_m(x, y)$ can be written down at once [Merriman, loc. cit.], but asymptotic formulas for those functions have never been obtained. Nevertheless, methods developed elsewhere [op. cit., §§6.6 and 6.7] together with the results of the present paper, enable us to establish certain results on the coefficients a_m in (58) and on the convergence of the series (58).

THEOREM 21. *In the situation of Theorem 20 case* ii), $p = 2$, *let the function* $u(x, y)$ *be single-valued and harmonic throughout the interior of* C_ρ, *but not throughout the interior of any* $C_{\rho'}$, $\rho' > \rho$. *Let the norm function* $n(x, y)$ *be positive and continuous on the boundary of* C. *Then the coefficients* a_m *in the formal expansion* (58) *of* $u(x, y)$ *satisfy the relation*

$$(59) \qquad \overline{\lim_{n \to \infty}} \, | a_m |^{1/m} = 1/\rho^{1/2}.$$

Reciprocally, if numbers a_m *are chosen arbitrarily so as to satisfy* (59), *the function* $u(x, y)$ *defined by* (58) *is single-valued and harmonic throughout the interior of* C_ρ, *but not throughout the interior of any* $C_{\rho'}$, $\rho' > \rho$.

It is of course possible throughout the present paper to consider degree of approximation not by harmonic polynomials of respective degrees n, but by harmonic polynomials $p_m(x, y)$ which are respectively linear combinations of the first $m + 1$ of the functions $1, r \cos \theta, r \sin \theta, r^2 \cos 2\theta, \cdots$. The results in the latter case are entirely analogous to the results in the former, and the results in either case may be readily established from the results in the other case. But in the latter case, the number ρ which now appears in such a relation as (10) is to be replaced by ρ^2.

THEOREM 22. *Under the hypothesis of the first part of Theorem* 21, *the series* (58) *can converge at every point of no region exterior to* C_ρ.

Under the conditions of Theorem 22, line segments of uniform convergence of the series (58) may be everywhere dense in the plane, as is illustrated by the example given in §6.

The conclusions of Theorems 21 and 22 apply also to the situations of Theorem 20, cases iii)–v), with $p = 2$. Details of the proof of Theorems 21 and 22 and of these extensions are left to the reader.

HARVARD UNIVERSITY.

Annals of Mathematics
Vol 38, No. 2, April, 1937

TENSOR COÖRDINATES OF LINEAR SPACES

By Wallace Givens

(Received July 2, 1936)

1. **Introduction.** Homogeneous coördinates of linear subspaces of a projective space were defined by Grassman and have been studied by Severi,[1] Antonelli,[2] and others. These studies have not, however, employed the notation and methods of the tensor calculus. In this paper we define contravariant and covariant tensor coördinates of the linear subspaces of a projective space and derive *ab initio* the quadratic relations which they satisfy.

These coördinates enable us to give elegant algebraic expression to the geometric operation of perspection and section, where the center of perspection is a linear space of any number of dimensions. Two expressions are found for the coördinate tensor of the join and intersection of two spaces and in verifying the equivalence of these two expressions we are led to a useful identity.

In §5 we introduce a non-singular quadratic form into the geometry and employ the matrix of this form to lower and raise tensor indices in the familiar fashion. We then derive some of the properties of the linear spaces which lie on a quadric. In particular, we find the algebraic characterization of the two families of rulings on a quadric in a complex projective space of $2\nu - 1$ dimensions and obtain the intersection properties of these rulings. Certain exceptional features of the quadric in 3-space are noted which imply special properties of the orthogonal group on four variables.

As a further application of the tensor coördinates we obtain a correspondence between tensor sets in a complex projective space of $2\nu - 1$ dimensions, $P_{2\nu-1}$, and collineations in a complex projective space of $2^\nu - 1$ dimensions, $P_{2^\nu-1}$. This correspondence is then employed to obtain a representation of the rulings on the quadric by means of points in two non-intersecting spaces of $2^{\nu-1} - 1$ dimensions in $P_{2^\nu-1}$. For $\nu = 3$ this leads to the well known Plücker-Klein representation of the points of a quadric in P_5 by the lines of a 3-space.

In §15 we obtain a matrix representation of the group of proper orthogonal matrices of order 2ν by matrices of order $2^{\nu-1}$ for values of ν greater than two. For odd values of ν this representation is shown to be (1-2) and for even values of ν it is shown to be (2-2). In §16 we give the corresponding theorems for the groups of proper orthogonal matrices of orders two and four.

The author desires to express his sincere gratitude to Professor O. Veblen for the stimulus of many discussions on the subject matter of this paper.

---

[1] *Annali di Matematica* 24 (3), 1915, pp. 89–120.
[2] *Annali della R. Scuola normale Sup. di Pisa*, Bd. III, 1883, pp. 71–77.

2. Definition of the Coördinate Tensors. An $(r-1)$-dimensional linear subspace, V, of the real or complex projective space P_{n-1} is determined by any r linearly independent points which it contains. If

$$A_1^i, A_2^i, A_3^i, \cdots, A_r^i \qquad (i = 1, 2, \cdots, n)$$

are coördinates of r such points, then the tensor

$$(2.1)\qquad \begin{aligned} V^{i_1 i_2 \cdots i_r} &= \frac{1}{r!}\,\delta^{i_1 i_2 \cdots i_r}_{s_1 s_2 \cdots s_r} \quad A_1^{s_1} A_2^{s_2} \cdots A_r^{s_r} \\ &= A_1^{[i_1} A_2^{i_2} \cdots A_r^{i_r]} \end{aligned}$$

where $\delta^{i_1 i_2 \cdots i_r}_{s_1 s_2 \cdots s_r}$ is the generalized Kronecker delta,[3] is said to be a contravariant coördinate tensor of V. Thus the coördinate tensor of the line determined by points A^i and B^i is $V^{ij} = \frac{1}{2}(A^iB^j - B^iA^j)$, and the components of the coördinate tensor of an $(r-1)$-dimensional space are $(1/r!)$ times the minors of the matrix of r rows and n columns formed by the coördinates of r independent points of the space.

In computations involving many indices we replace a set of indices such as $i_1 i_2 \cdots i_r$ by a single index in parentheses (i). The number of indices in a set (i) is denoted by $|\,(i)\,|$ or $|\,i\,|$, in words, the "length" of (i). Thus (2.1) becomes in this notation

$$(2.2)\qquad V^{(i)} = \frac{1}{|\,i\,|!}\,\delta^{(i)}_{(s)}\, A_1^{s_1} A_2^{s_2} \cdots A_{|s|}^{s_{|s|}},$$

where $|\,s\,| = |\,i\,| = r$. We shall also use symbols such as (k_i) to denote sets of ordinary indices $k_{i1}k_{i2} \cdots k_{ir}$, where $r = |\,k_i\,|$.

If we define $\epsilon_{(i)}$, $|\,i\,| = n$, to be $+1$ if (i) is an even permutation of

$$(1, 2, \cdots, n),$$

-1 if it is an odd permutation, and zero otherwise, and let $\epsilon_{(i)}$ be a covariant tensor of weight -1, then the components of $\epsilon_{(i)}$ have the same values in all coördinate systems. This numerically invariant tensor is always present in our geometry and we employ it to define a covariant tensor

$$(2.3)\qquad V_{(\bar{j})} = \frac{\rho(|\,i\,|)}{|\,i\,|!}\,\epsilon_{(j)\,(i)}\, V^{(i)}$$

associated with $V^{(i)}$. In this definition ρ is assumed to be a scalar function of the number of indices in the set (i) of weight $+1$ so that $V_{(\bar{j})}$ and $V^{(i)}$ have the same weight, which we shall take to be zero. For many purposes in projective geometry it would be sufficient to take $\rho = 1$ for all values of $|\,i\,|$ in an arbitrarily chosen coördinate system. When we introduce the matrix of a quadratic form we will determine the value of $\rho(\,|\,i\,|\,)$ so as to give a convenient calculus.

---

[3] For the definition and properties of these symbols see Chapter I of the Cambridge Tract No. 24, *Invariants of Quadratic Differential Forms*, by O. Veblen.

Observing that equations (2.3) essentially amount to no more than a renumbering of the components of $V^{(i)}$, we may invert the equations to get

$$V^{(i)} = \frac{1}{\rho(|i|)\,|j|!} V_{(\bar{j})} \epsilon^{(j)(i)}, \tag{2.4}$$

where $\epsilon^{(i)}$ is the numerically invariant contravariant tensor of weight $+1$ defined by the equations $\epsilon^{(i)} = \epsilon_{(i)}$. In (2.3) and (2.4) we have given rules for lowering and raising sets of skew-symmetric indices, the bar being added or removed by the operation. We also use (2.4) to define $W^{(\bar{i})}$ in terms of $W_{(j)}$ and then (2.3) gives $W_{(j)}$ in terms of $W^{(\bar{i})}$. So far we have no way of defining a tensor $V_{(i)}$ associated with $V^{(i)}$. In this paper we shall have no occasion to raise or lower part of the indices of a tensor by means of $\epsilon^{(i)}$ or $\epsilon_{(i)}$. We include this possibility in our calculus, however, by extending the bar over only those indices which have been lowered or raised.

The operations of raising and lowering sets of skew-symmetric indices give algebraic expression to the law of duality as applied to linear spaces. Thus if we choose coördinates so that $A_a^i = \delta_a^i (a = 1, 2, \cdots, r)$ in (2.2), $V_{(\bar{j})}$ will vanish unless (j) is a permutation of $((r + 1), (r + 2), \cdots, n)$ and consequently $V_{(\bar{j})}$ is a multiple of $\delta_{[j_1}^{r+1} \delta_{j_2}^{r+2} \cdots \delta_{j_{n-r}]}^{n}$. Hence a set of r independent points in V determines $V^{(i)}$ to within a factor and a set of $n - r$ linearly independent hyperplanes containing V determines $V_{(\bar{j})}$ to within a factor. A set of points $B_a^i = \lambda_a^b A_b^i$ will determine the coördinate tensor $\lambda V^{(i)}$, where λ is the determinant of the matrix $\| \lambda_a^b \|$. Consequently any two sets of r independent points in V determine the same pencil of tensors $\rho V^{(i)}$.

Also, the coördinates X^i of a point in V satisfy the linear equations

$$V_{\overline{(k)s}} X^s = 0, \tag{2.5}$$

and the coördinates of a hyperplane containing V satisfy the equations

$$V^{(k)s} X_s = 0. \tag{2.6}$$

Moreover, writing (2.5) in the form $\epsilon_{(i)(k)s} A_1^{k_1} A_2^{k_2} \cdots A_r^{k_r} X^s = 0$, we see that every solution of (2.5) is linearly dependent on $A_1^i, A_2^i, \cdots$, and A_r^i and so determines a point of V. A similar statement holds for (2.6). This completes the proof of the theorem

THEOREM (2.1). *A linear space determines its coördinate tensors to within a factor and is uniquely determined by them.*

3. The Quadratic Identities.

The coördinate tensors of a linear space are necessarily skew-symmetric and non-vanishing. Not all such tensors, however, are expressible in the form (2.1) and consequently a tensor $X^{(i)} = X^{[(i)]} \neq 0$ must satisfy additional conditions in order to be a coördinate tensor. One form of these conditions is given in the

THEOREM (3.1). *A necessary and sufficient condition that $V^{[(i)]} = V^{(i)} \neq 0$ shall be the coördinate tensor of a linear space is that*

$$\text{rank} \, \| V^{(k)s} \| = | (k)s |,$$

where the set of indices (k) *numbers the rows of the matrix and the single index* s *numbers the columns.* (Thus $\| V^{(k)s} \|$ is rectangular with $n^{|k|}$ rows and n columns.)

The necessity of the condition follows from observation that equations (2.6) have exactly $n - r$ independent solutions so that rank $\| V^{(k)s} \| = r = | (k)s |$. To prove the sufficiency of the condition we need only choose coördinates so that a set of $n - r$ linearly independent solutions of (2.6) are $\delta_i^{r+1}, \delta_i^{r+2}, \cdots$, and δ_i^n; then $V^{(i)}$ will be a multiple of $\delta_1^{[i_1} \delta_2^{i_2} \cdots \delta_r^{i_r]}$.

LEMMA 1. *If* $V^{(i)} = V^{[(i)]} \neq 0$, *then*

$$\text{rank} \, \| V^{(k)s} \| \geqq | (k)s |.$$

If rank $\| V^{(k)s} \|$ were less than $| (k)s | = r$, equations (2.6) would have more thant $n - r$ independent solutions. By a suitable choice of coördinate system a linearly independent set of these solutions could be taken to be $\delta_i^1, \delta_i^2, \cdots, \delta_i^t$, where $t > n - r$. Then $V^{(k)1} = V^{(k)2} = V^{(k)3} = \cdots = V^{(k)t} = 0$ and since r indices (i) cannot be chosen different from one another and from all the numbers $(1, 2, 3, \cdots, t)$, we would conclude from the skew-symmetry of $V^{(i)}$ that $V^{(i)} = 0$, contrary to hypothesis.

THEOREM (3.2). *If*

$$X^{(i)s} Y_{(j)s} = 0, \quad \text{where } X^{(x)} = X^{[(x)]} \neq 0, \quad Y_{(y)} = Y_{[(y)]} \neq 0,$$

and the indices range from one to n, then

(1) $| x | + | y | \leqq n$, *and*

(2) *if* $| x | + | y | = n$, *then* $Y_{(y)} = \lambda X_{(\bar{y})}$ *and* $X^{(x)}$ *and* $Y_{(y)}$ *are coördinate tensors of the same linear space.*

For each choice of the indices (j) the vector $Y_{(j)s}$ belongs to the right null-space of the matrix $\| X^{(i)s} \|$ and consequently

$$\text{rank} \, \| Y_{(j)s} \| \leqq n - \text{rank} \, \| X^{(i)s} \|. \tag{3.1}$$

But by Lemma 1,

$$| x | \leqq \text{rank} \, \| X^{(i)s} \| \text{ and } | y | \leqq \text{rank} \, \| Y_{(j)s} \|. \tag{3.2}$$

Combining these inequalities we have the first part of the theorem. Moreover, if $| x | + | y | = n$, the equality signs must hold in (3.2) so that by Theorem (3.1) $X^{(x)}$ is a coördinate tensor. Choosing coördinates so that

$$X^{(x)} = \lambda \delta_1^{[x_1} \delta_2^{x_2} \cdots \delta_{|x|}^{x_{|x|}]},$$

the hypothesis of the theorem implies that $Y_{(j)1} = Y_{(j)2} = \cdots = Y_{(j)\,|x|} = 0$ and consequently $Y_{(y)} = \mu\delta^{|x|+1}_{[y_1}\,\delta^{|x|+2}_{y_2}\,\cdots\,\delta^{n}_{y_{|y|}]}$ is a multiple of $X_{\overline{(y)}}$.

Theorem (3.3). *A necessary and sufficient condition that* $V^{(i)} = V^{[(i)]} \neq 0$ *shall be a coördinate tensor is that*

$$V^{(k)s}V_{\overline{(j)s}} = 0. \tag{3.3}$$

The sufficiency of the condition follows from Theorem (3.2) by putting $X^{(i)} = V^{(i)}$ and $Y_{(j)} = V_{\overline{(j)}}$. The necessity follows from the observation that when $V^{(i)}$ is expressible in the form (2), then for each set of values of the indices (k) and (j), $V^{(k)s}V_{\overline{(j)s}}$ is a multiple of a determinant in which two rows are equal.

Equations (3.3) obviously imply the vanishing of $V^{(k)(s)}V_{\overline{(j)(s)}}$ if (s) contains more than one index. In Theorem (5.4) we shall prove that

$$V^{(k)rs}V_{\overline{(j)rs}} = 0$$

implies the apparently stronger condition (3.3). This will prove the

Theorem (3.4). *A necessary and sufficient condition that* $V^{(i)} = V^{[(i)]} \neq 0$ *shall be a coördinate tensor is that*

$$V^{(k)rs}V_{\overline{(j)rs}} = 0. \tag{3.4}$$

The quadratic relations satisfied by the components of a coördinate tensor are usually given in the non-tensorial form[4]

$$V^{(m)ab}V^{(m)cd} + V^{(m)ac}V^{(m)db} + V^{(m)ad}V^{(m)bc} = 0, \tag{3.5}$$

where the set of indices (m) occurs twice but without being summed. These equations are obtained by multiplying (3.3) by $\epsilon^{(j)(m)bcd}$, summing, and putting $(m)a$ for (k). We shall here employ only the tensor form of the quadratic relations, (3.3) or (3.4).

4. **Joins and Intersections.** We call the linear space of smallest dimensionality which contains two linear spaces, X and Y, the join of X and Y and denote it by $X + Y$. Similarly, we call the space of greatest dimensionality contained in both X and Y the intersection (meet) of X and Y and denote it by XY. We write 0 and 1 for the null-set and the entire space, respectively. The coördinate tensors for 0 are a scalar and $e_{(i)}$. For 1 they are $\epsilon^{(i)}$ and a scalar.

Theorem (4.1). *If X, Y and $J = X + Y$ are linear spaces with the respective coördinate tensors $X^{(x)}$, $Y^{(y)}$, and $J^{(j)}$, and $XY = 0$, then $J_{\overline{(i)}}$ is proportional to $Y_{\overline{(i)(x)}}X^{(x)}$, and to $X_{\overline{(i)(y)}}Y^{(y)}$. If $XY \neq 0$, and $|x| + |y| \geqq n$, $Y_{\overline{(i)(x)}}X^{(x)} = X_{\overline{(i)(y)}}Y^{(y)} = 0$.*

[4] Bertini, *Einführung in die Projektive Geometrie Mehrdimensionaler Räume*, Vienna, 1924, p. 43, formula (14).

If X is the join of the $|x|$ points $A_1, A_2, \cdots, A_{|x|}$ and Y is the join of the $|y|$ points $B_1, B_2, \cdots, B_{|y|}$, then

$$J = A_1 + A_2 + \cdots + A_{|x|} + B_1 + B_2 + \cdots + B_{|y|}, \text{ and } XY = 0$$

implies that these $|x| + |y|$ points are linearly independent. Consequently,

$$J_{\overline{(i)}} = \epsilon_{(i)(x)(y)} A_1^{x_1} A_2^{x_2} \cdots A_{|x|}^{x_{|x|}} B_1^{y_1} B_2^{y_2} \cdots B_{|y|}^{y_{|y|}}$$

is different from zero and is by definition a covariant coördinate tensor of J. Using the definition of $X^{(x)}$ and $Y^{(y)}$,

$$J_{\overline{(i)}} = \epsilon_{(i)(x)(y)} X^{(x)} Y^{(y)} = (-1)^{|x||y|} \epsilon_{(i)(y)(x)} Y^{(y)} X^{(x)},$$

so that $J_{\overline{(i)}}$ is proportional to $Y_{\overline{(i)(x)}} X^{(x)}$ and to $X_{\overline{(i)(y)}} Y^{(y)}$. If $XY \neq 0$ the points $A_1, A_2, \cdots, A_{|x|}, B_1, B_2, \cdots, B_{|y|}$ are linearly dependent and the expressions vanish.

A similar proof gives the dual theorem.

THEOREM (4.2). *If $X + Y = 1$ and $XY = I$, then $I^{(i)}$ is proportional to $X^{(i)(r)} Y_{\overline{(r)}}$ and to $Y^{(i)(s)} X_{\overline{(s)}}$. If $X + Y \neq 1$, and $|x| + |y| \geqq n$, then*

$$X^{(i)(r)} Y_{\overline{(r)}} = Y^{(i)(s)} X_{\overline{(s)}} = 0.$$

THEOREM (4.3). *If X and Y are linear spaces with coördinate tensors $X^{(x)}$ and $Y^{(y)}$, respectively, then for each value of $|s| \leqq |y|$ and $\leqq n - |x|$ the "projection operator" $p_{(j)}^{(i)} = Y^{(i)(s)} X_{\overline{(j)(s)}}$ defines a transformation*

$$\varphi^{(i)} = p_{(j)}^{(i)} \psi^{(j)} \tag{4.1}$$

of linear subspaces of dimension $n - |x| - |s| - 1$ into linear subspaces of dimension $|y| - |s| - 1$. The transformation has the following properties:

1. *If $\psi^{(j)}$ is the coördinate tensor of a linear space ψ such that $X\psi = 0$ and $Y + X + \psi = 1$, then $\varphi^{(i)} (\neq 0)$ is a coördinate tensor of the linear space $Y(X + \psi)$.*
2. *If either or both of the conditions $X\psi = 0$ and $Y + X + \psi = 1$ fail to hold, then $\varphi^{(i)} = 0$.*

The theorem follows immediately by applying Theorem (4.1) to the spaces X and ψ and then Theorem (4.2) to the spaces Y and $(X + \psi)$. Dualizing Theorem (4.3) gives

THEOREM (4.4). *$\psi_{(i)} p_{(j)}^{(i)}$ is a coördinate tensor of $\psi Y + X$ unless $\psi + Y \neq 1$ or $\psi\, YX \neq 0$ and then it is zero.*

THEOREM (4.5). *If $X^{(x)}$ and $Y^{(y)}$ are coördinate tensors of linear spaces and we form the sequence of tensors*

$$Y^{(i)} X_{\overline{(j)}},\ Y^{(i_1)s_1} X_{\overline{(j_1)s_1}},\ Y^{(i_2)(s_2)} X_{\overline{(j_2)(s_2)}},\ \cdots$$
$$\cdots,\ Y^{(i_a)(s_a)} X_{\overline{(s_a)}} \quad \text{or} \quad Y^{(s_a)} X_{\overline{(j_a)(s_a)}}, \tag{4.2}$$

where $|s_r| = r$ *and* a *is the smaller of the two numbers* $n - |x|$ *and* $|y|$, *then the last non-vanishing tensor factors into the form*

$$I^{(i_r)} J_{(i_r)} \tag{4.3}$$

where $I^{(i_r)}$ *and* $J_{(i_r)}$ *are coördinate tensors of* $I = XY$ *and* $J = X + Y$, *respectively.*

The spaces X and Y intersect in a space of dimensionality $\leqq |x| - 1$ and $\leqq |y| - 1$. Hence if $I^{(i)}$ is a coördinate tensor of $I = XY$ we may put $|i| = |y| - r$ where $r \geqq 0$. Now applying Theorem (4.3), the expression

$$\varphi^{(i_r)} = Y^{(i_r)(s_r)} X_{\overline{(j_r)(s_r)}} \psi^{(j_r)} \tag{4.4}$$

is either zero or is the coördinate tensor of $\varphi = Y(X + \psi)$. In the latter case φ includes XY and its dimension, $|i_r| - 1 = |y| - r - 1$, equals the dimension of XY. Hence $\varphi = XY$ and $\varphi^{(i_r)} = I^{(i_r)}$. Under our assumptions a space ψ can be found satisfying the conditions $X\psi = 0$ and $Y + X + \psi = 1$ and for this space $\varphi^{(i_r)} \neq 0$ by Theorem (4.3).

It is now convenient to regard (4.4) as a vector equation $\varphi = M\psi$ where M is the rectangular matrix $\| Y^{(i_r)(s_r)} X_{\overline{(j_r)(s_r)}} \|$. Choosing as a basis for the set of all vectors $\psi^{(k)}$ the coördinate tensors $\delta^{(k)}_{(p)}$, where $|k| = |j_r|$ and (p) is any set of $|k|$ numbers between one and n, the matrix M transforms every vector of the basis either into the zero vector or into a multiple of $I^{(i_r)}$. Since M is not zero, it is of rank one and therefore factors into the product of two vectors. That is,

$$Y^{(i_r)(s_r)} X_{\overline{(j_r)(s_r)}} = I^{(i_r)} J_{(i_r)} \tag{4.5}$$

where $I^{(i_r)}$ is a coördinate tensor of XY. Applying Theorem (4.4) to this projection operator proves that $J_{(i_r)}$ is a coördinate tensor of $X + Y$.

Summing off $r + 1$ indices of $Y^{(y)}$ against $r + 1$ indices of $X_{\bar{(j)}}$ gives zero since each component of this tensor is a sum of terms of the form $P^i Q_i$, where P is a point of $I = XY$ and Q is a hyperplane containing $J = X + Y$ and *a fortiori* containing P. It is obvious that if any one of the terms of (4.2) vanishes, then all following terms also vanish.

If we exchange X and Y in (4.2) we get the sequence

$$\begin{gathered} X^{(i)} Y_{\bar{(j)}},\ X^{(i_1)s_1} Y_{\overline{(j_1)s_1}},\quad X^{(i_2)(s_2)} Y_{\overline{(j_2)(s_2)}},\ \cdots \\ \cdots,\ X^{(i_a)(s_a)} Y_{\bar{(s_a)}} \quad \text{or} \quad X^{(s_a)} Y_{\overline{(j_a)(s_a)}}, \end{gathered} \tag{4.6}$$

where $|s_r| = r$ and a is the smaller of the two numbers $n - |y|$ and $|x|$. It is to be observed that the number of terms in the sequences (4.2) and (4.6) are, respectively $n - |x| + 1$ and $n - |y| + 1$ if $n \leqq |x| + |y|$, and $|y| + 1$ and $|x| + 1$ if $n \geqq |x| + |y|$. Hence the sequences are in general of unequal length. For coördinate tensors $X^{(x)}$ and $Y^{(y)}$, Theorem (4.5) implies the

THEOREM (4.6). *If* $X^{(x)}$ *and* $Y^{(y)}$ *are any two skew-symmetric tensors, then the last non-vanishing terms in* (4.2) *and* (4.6) *are proportional.*

To prove this theorem without restricting $X^{(x)}$ and $Y^{(y)}$ to be coördinate tensors we use the definition (2.3) and its inverse (2.4) to express $Y_{\overline{(j)}}$ and $X^{(x)}$ in terms of $Y^{(y)}$ and $X_{\overline{(m)}}$, respectively. This gives

$$\begin{aligned} X^{(i)(r)}Y_{\overline{(j)(r)}} &= \lambda\delta^{(i)(m)}_{(y)(j)}\, Y^{(y)}X_{\overline{(m)}} \\ &= \lambda_1 Y^{(i)(s)}X_{\overline{(j)(s)}} + \lambda_2\delta^{[i_1}_{[j_1}\, Y^{i_2 \cdots i_{|i|}](s_1)}X_{j_2 \cdots j_{|j|}](s_1)} + \end{aligned} \tag{4.7}$$

$+$ additional terms in which more than $|s| + 1$ indices are summed, where $|s_1| = |s| + 1$ and λ, λ_1, and λ_2 are scalars which depend on the lengths of the various indices. Consequently, if $Y^{(i)(s)}X_{\overline{(j)(s)}}$ is the last non-vanishing term in (4.2),

$$x^{(i)(r)}Y_{\overline{(j)(r)}} = \lambda_1 Y^{(i)(s)}X_{\overline{(j)(s)}}, \tag{4.8}$$

where $|r| = |s| - |y| + |x|$, and hence $X^{(i)(r)}Y_{\overline{(j)(r)}}$ must be the last non-vanishing term in (4.6).

5. The Quadratic Form. We now introduce into our geometry a fundamental quadratic form $\gamma_{ij}X^iX^j$ and use the symmetric matrix $\|\gamma_{ij}\|$ to lower single tensor indices by the familiar rule, $X_i = \gamma_{ij}X^j$. Covariant indices are raised by the rule $X_i\gamma^{ij} = X^j$, where $\gamma^{ki}\gamma_{kj} = \delta^i_j$.

If we lower each of the indices of a skew-symmetric tensor $X^{(j)}$ by γ_{ij} the resulting tensor is skew-symmetric, and is denoted by $X_{(j)}$. Applying the rule (2.4) to raise the indices (j) by means of $\dfrac{1}{\rho(|i|)\,|j|!}\,\epsilon^{(j)(i)}$ gives the tensor $X^{(i)}_{-}$. Alternatively, we could first lower the indices of $X^{(j)}$ by the rule (2.3) to get $X_{\overline{(i)}}$ and then raise them with γ^{ij}. The result of these two processes is the same to within a factor. Indeed, by using (2.3), (2.4), and the formula for the determinant of the matrix $\|\gamma^{ij}\|$, we find that

$$\gamma^{(s)(i)}X_{\overline{(s)}} = [\rho(|i|)\,\rho(n - |i|)\,\gamma^{-1}(-1)^{|i|(n-|i|)}]\,X^{(i)}_{-}, \tag{5.1}$$

where $\gamma^{(s)(i)} = \gamma^{s_1i_1}\gamma^{s_2i_2}\cdots\gamma^{s_{|i|}i_{|i|}}$ and γ is the determinant $|\gamma_{ij}|$.

The scalar factor in the right member of (5.1) will be $+1$ if we put

$$\rho(|i|) = \gamma^{\frac{1}{2}}(-1)^{\frac{1}{2}|i|(n-|i|)}. \tag{5.2}$$

This choice has the disadvantage that $\rho(|i|)$ is sometimes imaginary. This cannot always be avoided without introducing additional complications into the calculus for in the special case in which $n = 2\nu$ and $|i| = \nu$, (5.1) reduces to

$$\gamma^{(s)(i)}X_{\overline{(s)}} = \rho^2(\nu)\gamma^{-1}(-1)^{\nu}X^{(i)}_{-} \tag{5.3}$$

so that $\rho(\nu) = \gamma^{\frac{1}{2}}(-1)^{\frac{1}{2}\nu}$. However, if $\|\gamma_{ij}\|$ is a real matrix of signature $(+-+-\cdots+-)$, then $\rho(\nu)$ is real (cf. Theorem (6.2)).

We now restate our rules for lowering and raising sets of skew-symmetric indices. They are

$$X^{-}_{j} = \frac{\{i\}\,\gamma^{\frac{1}{2}}}{|\,i\,|\,!}\,\epsilon_{(j)(i)}\,X^{(i)},$$

(5.4) and

$$X^{i}_{-} = \frac{\{j\}^{-1}\,\gamma^{-\frac{1}{2}}}{|\,j\,|\,!}\,X_{(j)}\,\epsilon^{(j)(i)},$$

where

$$\{i\} = (-1)^{\frac{1}{2}|\,i\,|(n-|\,i\,|)}. \tag{5.5}$$

In (5.4) if a bar ($^{-}$) is added to the X in the right member it is erased from the left member. We observe that $\{i\} = \{k\}$ if $|\,i\,| + |\,k\,| = n$. The operations of our calculus are summarized in the theorem[5]

THEOREM (5.1). *If single tensor indices are lowered and raised by γ_{ij} and γ^{ij} while sets of skew-symmetric indices are lowered and raised by the rules (5.4), the resulting calculus is consistent and gives the relations indicated in the diagram*

$$\begin{tikzcd}
X^{(i)} \arrow[rr, leftrightarrow, "\epsilon"] \arrow[d, "\gamma"'] & & X^{-}_{(m)} \arrow[d, leftrightarrow, "\gamma"] \\
X_{(i)} \arrow[rr, leftrightarrow, "\epsilon"'] & & X^{(m)}_{-}
\end{tikzcd} \tag{5.6}$$

The rules regulating the manner in which sets of summed indices may be raised and lowered are computed to be

$$X^{(i)}Y_{(i)} = X_{(i)}Y^{(i)}, \qquad X^{(i)}_{-}Y^{-}_{(i)} = X^{-}_{(i)}Y^{(i)}_{-}, \tag{5.7}$$

and

$$\frac{1}{|\,i\,|\,!}\,X^{(i)}\,Y_{(i)} = \frac{\{i\}^{2}}{|\,k\,|\,!}\,X^{-}_{(k)}\,Y^{(k)}_{-}. \tag{5.8}$$

The numerical coefficients in equations (4.7) may be evaluated by computation with (5.4). The result of this computation for the first of these coefficients is contained in the

---

[5] If $\gamma_{ij} = -\gamma_{ji}$, a factor $(-1)^{|\,i\,|}$ must be added to the right member of (5.1). Theorem (5.1) will continue to hold if we put $\{i\} = (-1)^{\frac{1}{2}|\,i\,|(n-|\,i\,|+1)}$ in (5.4). In order that $|\,\gamma_{ij}\,| \neq 0$, n must be even and hence $\{i\}$ is always real. Equations (5.7) of course require modification when $\gamma_{ij} = -\gamma_{ji}$.

Theorem (5.2). *If $X^{(x)}$ and $Y^{(y)}$ are skew-symmetric tensors, $|i|$ is $\leqq |x|$ and $\leqq |y|$, and $|j|$ is $\leqq n - |x|$ and $\leqq n - |y|$, then*

$$
\frac{\{r\}}{|r|!} X^{(i)(r)} Y_{\overline{(j)(r)}} = (-1)^{|r||s|} \frac{\{s\}}{|s|!} Y^{(i)(s)} X_{\overline{(j)(\bar{s})}} + \sum_{|v|=|s|+1}^{M} \rho_{|v|} \delta^{[(i_1)}_{[(j_1)} Y^{(i_2)](v)} X_{\overline{(j_2)](v)}}, \tag{5.9}
$$

where $(i) = (i_1)(i_2)$, $(j) = (j_1)(j_2)$, and M is the smaller of the two numbers $|y|$ and $n - |x|$.

This useful identity depends only upon the rules for lowering and raising sets of skew-symmetric indices by means of the numerical tensors and does not involve the tensor γ_{ij}. If $|\gamma_{ij}|$ is not available we suppose the factor γ in (5.4) to be an arbitrary scalar of weight two.

Special cases of (5.9) are stated in the following theorems.

Theorem (5.3). *If $X^{(x)}$ and $Y^{(y)}$ are skew-symmetric and $|x| + |y| = n$, then*

$$
\frac{1}{|r|!} X^{(i)(r)} Y_{\overline{(j)(r)}} = (-1)^{|x||y|+|i|} \frac{1}{|s|!} Y^{(i)(s)} X_{\overline{(j)(s)}} \tag{5.10}
$$

+ additional terms in which more than $|s|$ indices are summed.

Theorem (5.4). *If $V^{(v)r(s)} V_{\overline{(w)r(s)}} = 0$ and $|s|$ is odd, then $V^{(v)m(s)} V_{\overline{(w)p(s)}} = 0$.*

To prove this we put $X^{(i)} = Y^{(i)} = V^{(i)}$ in (5.9) and replace (i) and (j) by $(v)r$ and $(w)p$, respectively. This theorem, for $|s| = 1$, was used in the proof of Theorem (3.4).

Theorem (5.5). *If $X^{(x)}$ and $Y^{(y)}$ are skew-symmetric and $|x| + |y| = n$, then*

$$
\frac{1}{|r|!} X^{i(r)} Y_{\overline{j(r)}} + (-1)^{|x||y|} \frac{1}{|s|!} Y^{i(s)} X_{\overline{j(s)}} = \frac{(X^{(x)} Y_{\overline{(x)}})}{|x|!} \delta^i_j. \tag{5.11}
$$

This is the special case of (5.10) in which all but one of the indices of $X^{(x)}$ are summed against indices of $Y_{\overline{(x)}}$. The scalar factor in the right member of (5.11) is most readily obtained by taking the trace of the left member and using (5.8).

An application of our calculus to an elementary problem in projective geometry is contained in the

Theorem (5.6). *If $X^{(x)}$ and $Y^{(y)}$ are coördinate tensors of linear spaces, X and Y, where $X + Y = 1$ and $XY = 0$, then*

$$
S^i_j = \frac{1}{|r|!} X^{i(r)} Y_{\overline{j(r)}} - (-1)^{|x||y|} \frac{1}{|s|!} Y^{i(s)} X_{\overline{j(s)}} \tag{5.12}
$$

defines a collineation $\varphi^i = S^i_j \psi^j$ with the pointwise invariant spaces X and Y. This collineation is of period two unless X is 0 or 1 and then it is the identity.

In proving this theorem it is convenient to abbreviate (5.12) to

(5.13) $$S = X \cdot Y - \lambda Y \cdot X,$$

and by a similar abbreviation to replace the identity (5.11) by

(5.14) $$X \cdot Y + \lambda Y \cdot X = \mu 1.$$

Then S can be written in the forms

(5.15) $$S = 2X \cdot Y - \mu 1,$$

and

(5.16) $$S = \mu 1 - 2\lambda Y \cdot X.$$

The square of the collineation $\varphi^i = S^i_j \psi^j$ is the identity, for

$$\begin{aligned} SS &= (2X \cdot Y - \mu 1)(\mu 1 - 2\lambda Y \cdot X) \\ &= 2\mu(X \cdot Y + \lambda Y \cdot X) - \mu^2 1 \\ &= \mu^2 1, \end{aligned}$$

where the "product" $-4\,\lambda(X \cdot Y)(Y \cdot X)$ vanishes since it contains the tensor $Y_{\overline{j(r)}} Y^{j(s)}$ which is zero by Theorem (3.3). The collineation is non-singular since $\mu = 0$ would imply $XY \neq 0$, contrary to the hypothesis. A point ψ^i in the space Y satisfies the conditions $Y_{\overline{j(r)}} \psi^j = 0$, so that using (5.15) we have

$$S^i_j\, \psi^j = \left(\frac{2}{|r|!} X^{i(r)}\, Y_{\overline{j(r)}} - \mu \delta^i_j \right) \psi^j = -\mu \psi^i.$$

Hence the points of Y are invariant. Similarly, using (5.16), the points of X are invariant.

6. **Linear Spaces on a Quadric.** The tensors $V^{(i)}$ and $V_{\bar{(i)}}$ have been interpreted as coördinate tensors of the same linear space. If we write

$$V^{(i)} = \sum \pm A^{i_1} B^{i_2} \cdots H^{i_{|i|}},$$

we observe that

$$V_{(i)} = \sum \pm A_{i_1} A_{i_2} \cdots H_{i_{|i|}}.$$

The hyperplanes $A_i,\ B_i, \cdots, H_i$ are the polars, respectively, of the points $A^i, B^i, \cdots, H^i$ in the quadric

(6.1) $$Q:\ \ \gamma_{ij} X^i X^j = 0.$$

Hence $V_{(i)}$ is a covariant coördinate tensor of the polar of the space $V^{(i)}$. Of course $V^{\bar{(i)}}$ is a contravariant coördinate tensor of this same space. Since a linear space $V^{(i)}$ lies on Q if and only if it is contained in its polar space, we can use Theorem (4.5) to get the

THEOREM (6.1). *A space with coördinate tensor $V^{(v)}$ lies on Q if and only if*

$$V^{(i)s}V_{(j)s} = 0. \tag{6.2}$$

Reference to the first part of Theorem (3.2) shows that (6.2) implies $2 \mid v \mid \leqq n$, so that the maximum number of linearly independent points in a space on Q is $\frac{1}{2}(n - 1)$ when n is odd and $\frac{1}{2}n$ when n is even. We now restrict our considerations to the latter case so that (6.1) defines a quadric in a projective space, $P_{2\nu-1}$, of $2\nu - 1$ dimensions. A linear subspace of $P_{2\nu-1}$ of $\nu - 1$ dimensions will be called an "axis." An axis then has a coördinate tensor $V^{(i)}$, where $\mid i \mid = \nu$. The polar of an axis is again an axis so that $V^{(i)}$ will lie on the quadric if and only if it coincides with its polar space. That is,

$$V^{(i)} = \rho V^{(i)}_{-}. \tag{6.3}$$

Lowering the indices (i) with $\gamma_{(s)(i)}$ and raising them again with $\epsilon^{(s)(i)}$ gives $V^{(i)}_{-} = \rho V^{(i)}$ and consequently $\rho = \pm 1$.

THEOREM (6.2). *A necessary and sufficient condition that a coördinate tensor $V^{(i)}(\mid i \mid = \nu)$ shall determine an axis on the quadric* (6.1) *is that either*

$$F_+ \colon \quad V^{(i)} = +V^{(i)}_{-}, \tag{6.4}$$

or

$$F_- \colon \quad V^{(i)} = -V^{(i)}_{-}. \tag{6.5}$$

The axes on Q are thus separated into the two families F_+ and F_-, characterized by (6.4) and (6.5) respectively. The association of the plus sign with one family and the minus sign with the other is a matter of arbitrary choice. Indeed it is determined by the selection of one of the two values of the factor $\gamma^{\frac{1}{2}}$ occurring in (5.4).

THEOREM (6.3). *If V and W are two axes on Q, then*

$$\textit{dimension of } VW \equiv \nu - 1 \pmod 2 \tag{6.6}$$

if V and W belong to the same family, and

$$\textit{dimension of } VW \equiv \nu - 2 \pmod 2 \tag{6.7}$$

if V and W belong to different families.[6]

Since V and W are axes on Q,

$$V^{(i)} = \lambda_V V^{(i)}_{-}, \qquad W^{(i)} = \lambda_W W^{(i)}_{-}, \qquad \mid i \mid = \nu, \tag{6.8}$$

and the spaces will belong to the same family if $\lambda_V \lambda_W = +1$ and to different families if $\lambda_V \lambda_W = -1$. Let us put $D(VW) = 1 +$ dimension of VW. Then to prove the theorem it is sufficient to show that $D(VW) > \mu$ implies $D(VW) >$

[6] A proof of this theorem based upon stereographic projection is given by Bertini, *loc. cit.*, p. 143.

$\mu + 1$ if $\lambda_V \lambda_W = +1$ and $\nu - \mu \equiv 0 \pmod 2$, or if $\lambda_V \lambda_W = -1$ and $\nu - \mu \equiv 1 \pmod 2$.

Assuming that $D(VW) > \mu$ and applying Theorem (4.5), we have $V^{\underline{(i)(s)}} W_{\overline{(j)(s)}} = 0$ for $|s| \geqq \nu - \mu$. Hence if we put $X^{(v)} = V^{(v)}$ and $Y_{(\bar{v})} = W_{(v)}$ in equations (5.10) they reduce to

$$(6.9) \qquad V^{(i)(r)} W_{(j)(r)} = -(-1)^{\nu-\mu} W^{\underline{(i)(r)}} V_{\overline{(j)(r)}}, \text{ where } |i| = \mu + 1.$$

On account of (6.8) these equations imply

$$(6.10) \qquad V^{(i)(r)} W_{(j)(r)} = -(-1)^{\nu-\mu} \lambda_V \lambda_W W^{(i)(r)} V_{(j)(r)}.$$

We wish to prove that under certain conditions $W^{(i)(r)} V_{(j)(r)} = 0$, for this implies $D(VW) > |i| = \mu + 1$. If this is not the case, then by Theorem (4.5) $W^{(i)(r)} V_{(j)(r)} = I^{(i)} J_{(j)}$ where $I^{(i)}$ and $J_{(j)}$ are coördinate tensors of VW and $V + W$, respectively. Since $V + W$ is the polar of VW, $J_{(j)} = \rho \gamma_{(j)(s)} I^{(s)}$ and consequently $W^{(i)(r)} V^{(j)}{}_{(r)} = \rho I^{(i)} I^{(j)}{}_{(r)}$ is symmetric in the sets (i) and (j). Hence $W^{(i)(r)} V^{(j)}{}_{(r)} = W^{(j)(r)} V^{(i)}{}_{(r)} = V^{(i)(r)} W^{(j)}{}_{(r)}$. Substituting in (6.10) now gives

$$(6.11) \qquad (1 + \lambda_V \lambda_W (-1)^{\nu-\mu}) V^{(i)(r)} W_{(j)(r)} = 0,$$

and therefore $V^{(i)(r)} W_{(j)(r)} = 0$ if $\lambda_V \lambda_W = +1$ and $\nu - \mu \equiv 0 \pmod 2$ or if $\lambda_V \lambda_W = -1$ and $\nu - \mu \equiv 1 \pmod 2$.

THEOREM (6.4). *If $P_{2\nu-1}$ is complex, or if it is real and the signature of Q is zero, then Q contains axes. If $P_{2\nu-1}$ is real, Q contains axes only if its signature is zero.*

We omit the proof.

THEOREM (6.5). *If Q contains axes, then axes V and W on Q exist for which $D(VW) = \alpha$, where α is any integer $\geqq 0$ and $\leqq \nu$.*

Under the hypothesis of the theorem we may choose coördinates in the real or complex $P_{2\nu-1}$ so that the equation of Q is

$$(6.12) \qquad X^1 X^{\nu+1} + X^2 X^{\nu+2} + \cdots + X^\nu X^{2\nu} = 0.$$

Let us call the vertices of the coördinate 2ν-point $E_1, E_2, \cdots, E_\nu, E_1', E_2', \cdots,$ and E_ν', where the coördinates of $E_a (a = 1, 2, \cdots, \nu)$ are δ_a^i and of E_a' are $\delta_{a+\nu}^i$. Then the spaces

$$V = E_1 + E_2 + \cdots + E_\alpha + E_{\alpha+1} + \cdots + E_\nu$$

(6.13) and

$$W = E_1 + E_2 + \cdots + E_\alpha + E_{\alpha+1}' + \cdots + E_\nu'$$

lie on Q and $D(VW) = \alpha$.

THEOREM (6.6). *If the quadric, Q, in $P_{2\nu-1}$ contains the distinct axes V and W, $\nu > 2$, and β is any integer $\geqq 0$ and $\leqq \nu$, then there exists an axis, A, on Q such that $D(AV) = \beta \neq D(AW)$.*

The theorem cannot be extended to include the case in which $\nu = 2$, $\beta = 1$,

and $D(VW) = 0$ for then V and W would be non-intersecting lines on a quadric in P_3 and any line on Q which intersects one of two such lines in a point also intersects the other in a point. In the other cases which occur when $\nu = 1$ or 2 the theorem is trivially satisfied.

Let us put $D(VW) = \alpha\ (< \nu)$. Then it can be shown that by a suitable choice of coördinate system we may take the equation of Q to be (6.12) while V and W are the faces of the coördinate 2ν-point given by (6.13). We now distinguish several cases:

Case 1. $\beta < \alpha$. Take $A = E_1 + E_2 + \cdots + E_\beta + E'_{\beta+1} + \cdots + E'_\nu$. Then $D(AV) = \beta$, and $D(AW) = \beta + \nu - \alpha > \beta$.

Case 2. $\beta \geqq \alpha$ and $2\beta \neq \nu + \alpha$. Take A as in Case 1. Then $D(AV) = \beta$, and $D(AW) = \alpha + \nu - \beta \neq \beta$.

Case 3. $2\beta = \nu + \alpha$. Take $A = E'_1 + E'_2 + \cdots + E'_{\nu-\beta} + E_{\nu-\beta+1} + \cdots + E_\nu$. Then $D(AV) = \beta$, and if $\alpha + \beta \geqq \nu$, $D(AW) = \alpha + \beta - \nu \neq \beta$, or if $\alpha + \beta < \nu$, $D(AW) = \nu - \alpha - \beta$, which is different from β unless $2\beta = \nu - \alpha$. In this event $\alpha = 0$, and $2\beta = \nu$.

Case 4. $2\beta = \nu > 2, \alpha = 0$. The 3-space $E_{\nu-1} + E_\nu + E'_{\nu-1} + E'_\nu$ intersects Q in a non-degenerate quadric which contains the non-intersecting lines $E_{\nu-1} + E_\nu$ and $E'_{\nu-1} + E'_\nu$. Let $F + G$ be any third line of the regulus containing these two lines. If we put $A = E_1 + E_2 + \cdots + E_\beta + E'_{\beta+1} + \cdots + E'_{\nu-2} + F + G$, then $D(AV) = \beta = \frac{1}{2}\nu$ and $D(AW) = \beta - 2 \neq \beta$.

THEOREM (6.7). *If two non-singular collineations of $P_{2\nu-1}(\nu > 2)$ leave Q invariant and effect the same transformation of the axes of F_+, then they are identical.*

Distinct axes of F_+ are transformed by the collineations into distinct axes which again lie on Q. Since intersection properties are preserved, the transformed axes will all belong to the same family, which may be either F_+ or F_-. The following proof applies to both these cases.

By the preceding theorem an axis of F_- is unambiguously determined by the axes of F_+ which intersect it in a space of $\beta - 1$ dimensions, where β is any number (say $\nu - 1$) congruent to $\nu - 1 \pmod 2$ and between 0 and $\nu - 1$. Hence the two collineations must effect the same transformation of the axes of F_-. Since every point on Q is the complete intersection of two suitably chosen axes on Q (of the same or different families), the two collineations transform the points of Q in the same way and are therefore identical throughout the space.

7. Representation of Points of $P_{2\nu-1}$ as Involutoric Collineations of $P_{2^\nu-1}$. The non-singular quadratic form $\gamma_{ij}X^iX^j$ cannot be written as the square of a linear form in the 2ν indeterminates $X^1, X^2, \cdots, X^{2\nu}$ so long as the coefficients of the linear form are restricted to the field of complex numbers. The equation

$$(X^i\gamma_i)^2 = (\gamma_{ij}X^iX^j)\ 1 \tag{7.1}$$

is, however, an identity in X^i if the coefficients γ_i are suitably chosen matrices of order 2^ν. To show this we first introduce a coördinate system in $P_{2\nu-1}$ in which

$$\gamma_{ij}X^iX^j = (X^1)^2 + (X^2)^2 + \cdots + (X^{2\nu})^2. \tag{7.2}$$

We shall call a coördinate system cartesian if $\gamma_{ij}X^iX^j$ has this form in it. If we assume that γ_i behaves as a covariant vector under transformations of coördinates in $P_{2\nu-1}$, a solution of (7.1) in cartesian coördinates gives rise to a solution of (7.1) in arbitrary coördinates.

In a cartesian coördinate system, equating coefficients in (7.1) gives

$$(\gamma_i)^2 = 1, \quad \text{and} \quad \gamma_i\gamma_j = -\gamma_j\gamma_i \qquad (i \neq j). \tag{7.3}$$

For $\nu = 1$ matrices satisfying these equations are

$$\gamma_1^{(1)} = \left\| \begin{matrix} 0 & i \\ -i & 0 \end{matrix} \right\| \quad \text{and} \quad \gamma_2^{(1)} = \left\| \begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix} \right\|. \tag{7.4}$$

The recursion formulae

$$\begin{gathered} \gamma_a^{(\nu)} = \left\| \begin{matrix} \gamma_a^{(\nu-1)} & 0 \\ 0 & -\gamma_a^{(\nu-1)} \end{matrix} \right\| \qquad (a = 1, 2, \cdots, 2(\nu-1)), \\ \gamma_{2\nu-1}^{(\nu)} = \left\| \begin{matrix} 0 & i\,1_{2^{\nu-1}} \\ -i\,1_{2^{\nu-1}} & 0 \end{matrix} \right\|, \quad \text{and} \quad \gamma_{2\nu}^{(\nu)} = \left\| \begin{matrix} 0 & 1_{2^{\nu-1}} \\ 1_{2^{\nu-1}} & 0 \end{matrix} \right\|, \end{gathered} \tag{7.5}$$

then supply solutions of equations (7.3) for every value of ν.

The matrices $\gamma_i = \| \gamma_i{}^A{}_B \|$ $(i = 1, 2, \cdots, 2\nu; A, B = 1, 2, \cdots, 2^\nu)$ form a basis for a linear family of involutoric collineations in $P_{2^\nu-1}$. A coördinate vector X^i of a point in $P_{2\nu-1}$ determines a matrix $X = \| X^A{}_B \|$ by means of the equations

$$X = X^i\gamma_i. \tag{7.6}$$

Equally well, these equations determine a representation

$$\rho X^i \leftrightarrow \rho \, \| X^A{}_B \| \tag{7.7}$$

of the points of $P_{2\nu-1}$ by involutoric collineations of $P_{2^\nu-1}$ defined by the pencils $\rho \, \| X^A{}_B \|$ of matrices.

8. The Correspondence Between Tensor Sets and Matrices. In a cartesian coördinate system the product of any number of distinct matrices of the set $\gamma_1, \gamma_2, \cdots, \gamma_{2\nu}$ has zero trace. This follows from the equation

$$\gamma_v(\gamma_{i_1}\gamma_{i_2} \cdots \gamma_{i_r})\gamma_v^{-1} = -(\gamma_{i_1}\gamma_{i_2} \cdots \gamma_{i_r}), \tag{8.1}$$

which follows from (7.3) if v is equal to i_1 when r is even and different from $i_1, i_2, \cdots$, and i_r when r is odd.

By using the commutation rules (7.3) and the relations

$$\text{Trace } (\gamma_i\gamma_j \cdots \gamma_s) = 0, \qquad (i, j, \cdots, s \text{ all different}) \tag{8.2}$$

it can be shown that the matrices

$$1, \gamma_i, \gamma_i\gamma_j \ (i < j), \cdots, \gamma_{i_1}\gamma_{i_2}\gamma_{i_3} \cdots \gamma_{i_{2\nu}}, \qquad (i_1 < i_2 < \cdots < i_{2\nu}) \tag{8.3}$$

are linearly independent. Since there are $2^{2\nu}$ of them every matrix of order 2^ν can be uniquely expressed as a linear sum of these matrices. That is, an arbitrary matrix $X = \| X^A{}_B \|$ of order 2^ν is expressible in the form

$$X = \sum_{|i|=0}^{2\nu} \frac{1}{|i|!} X^{(i)} s_{(i)}, \tag{8.4}$$

where $X^{(i)} = X^{[(i)]}$, $s_{(i)} = s_{[(i)]}$ and for different values of the indices $(i) = i_1 i_2 \cdots i_{|i|}$, $s_{(i)}$ is defined by the equations

$$s_{(i)} = \gamma_{i_1}\gamma_{i_2} \cdots \gamma_{i_{|i|}}. \tag{8.5}$$

It is understood that when $|i| = 0$, $s_{(i)} = 1$. When X is a matrix of the linear family $X^i\gamma_i$ the coefficients $X^{(i)}$ in (8.4) vanish unless $|i| = 1$ so that equations (8.4) include (7.6) as a special case.

The quantity $s_{(i)} = \| s^A_{(i)B} \|$ was defined to be skew-symmetric in the indices (i). This property is preserved under transformations of coördinates. Indeed, we may define $s_{(i)}$ by the formula

$$s_{(i)} = \frac{1}{|j|!} \delta^{(j)}_{(i)} \gamma_{j_1} \gamma_{j_2} \cdots \gamma_{j_{|j|}}, \tag{8.6}$$

which reduces to (8.5) in a cartesian coördinate system.

We can solve for the coefficients $X^{(i)}$ in (8.4) in terms of the matrix X. To do this we employ cartesian coördinates to prove that Trace $s^{(i)}s_{(j)} = 0$ if $|i| \neq |j|$ and otherwise it is a multiple of $\delta^{(i)}_{(j)}$. The exact relation

$$\text{Trace } (s^{(i)}s_{(j)}) = (-1)^{\frac{1}{2}|i|(|i|-1)}\, 2^\nu\, \delta|{}^i_j|\, \delta^{(i)}_{(j)} \tag{8.7}$$

may be verified in a cartesian coördinate system and evidently remains valid under coördinate transformations. Multiplying (8.4) by $s^{(i)}$ and taking the trace therefore gives

$$X^{(i)} = (-1)^{-\frac{1}{2}|i|(|i|-1)}\, 2^{-\nu}\, \text{Trace } X\, s^{(i)} \tag{8.8}$$

and we have the

THEOREM (8.1). *Equations* (8.4) *and* (8.8) *establish a* (1-1) *correspondence*

$$\| X^A{}_B \| \leftrightarrow \{X, X^i, X^{ij} = -X^{ji}, \cdots, X^{(i)} = X^{[(i)]}(|i| = 2\nu)\} \tag{8.9}$$

between matrices of order 2^ν *and sets of skew-symmetric tensors.*[7]

9. The Collineations Corresponding to Linear Spaces on the Quadric.

The correspondence of this theorem allows us to represent a linear subspace of $P_{2\nu-1}$ as a collineation in the spin[8] space $P_{2^\nu-1}$. To do this it is only necessary to

[7] The ordinary rule for the multiplication of two matrices gives rise to a rule for the "multiplication" of the corresponding tensor sets. A discussion of this multiplication of tensor sets is included in the paper by R. Brauer and H. Weyl, *Spinors in n Dimensions*, Am. Jour. Math., LV II (2), 1935, pp. 425–449.

[8] The term "spin space" is appropriate for $P_{2^\nu-1}$ because for $\nu = 2$ it is the space of the spin variables introduced by Dirac in his theory of the electron.

take the tensor set to be $\{0, 0, \cdots, 0, A^{(i)}, 0, \cdots, 0\}$ where $A^{(i)}$ is a coördinate tensor of the linear subspace.

A lemma which we shall need in the proof of the next theorem is

LEMMA 2. *If A and B are square matrices, $A^2 = B^2 = 0$, and $AB = \pm BA$, then*

$$\text{rank } AB \leqq \tfrac{1}{2} \text{ rank } A. \tag{9.1}$$

This is obtained from the general relation[9] (holding between any three matrices A, B and C)

$$\text{rank } AB + \text{rank } BC \leqq \text{rank } B + \text{rank } ABC$$

by putting $C = A$.

THEOREM (9.1). *If $L^{(i)}$ is a coördinate tensor of a linear space L on the quadric and $\lambda = \dfrac{1}{|i|!} L^{(i)} s_{(i)}$, then rank $\lambda = 2^{\nu - |i|}$ and unless L is the null set $\lambda^2 = 0$.*

Let us take $L_1, L_2, \cdots, L_{|i|}$ to be $|i|$ linearly independent points on L. (If $|i| = 0$, L is the null set and $\lambda = 1$.) Then $L = L_1 + L_2 + \cdots + L_{|i|}$. It is possible to find additional points $L_{|i|+1}, \cdots, L_\nu$ such that $A = L_1 + \cdots + L_\nu$ is an axis on the quadric. Then $\gamma_{ij} L_a^i L_b^j = 0$ for $a, b = 1, 2, \cdots, \nu$, where L_a^i is a coördinate vector of the point L_a.

Equating coefficients of the arbitrary variables X^i in (7.1) gives

$$\tfrac{1}{2}(\gamma_i \gamma_j + \gamma_j \gamma_i) = \gamma_{ij} 1. \tag{9.2}$$

Multiplying both members by $L_a^i L_b^j$ and summing gives

$$\lambda_a \lambda_b = -\lambda_b \lambda_a \qquad (a, b = 1, \cdots, \nu) \tag{9.3}$$

where $\lambda_a = L_a^i \gamma_i$.

In particular, $(\lambda_a)^2 = 0$, so that rank $\lambda_a \leqq \frac{1}{2}$ (order of λ_a) $= 2^{\nu-1}$. The product of any number of the matrices λ_a also has square zero for $(\lambda_a \lambda_b \cdots \lambda_p)^2 = \pm (\lambda_a)^2 (\lambda_b)^2 \cdots (\lambda_p)^2 = 0$. Since $L^{(i)} = L_1^{[i_1} L_2^{i_2} \cdots L_{|i|}^{i_{|i|}]}$, $\lambda = \lambda_1 \lambda_2 \cdots \lambda_{|i|}$ and therefore $\lambda^2 = 0$.

Repeated application of the lemma now gives the chain of inequalities

$$\begin{aligned}
&\text{rank } \lambda_1 && \leqq 2^{\nu-1} && \\
&\text{rank } (\lambda_1 \lambda_2) && \leqq \tfrac{1}{2} \text{ rank } \lambda_1 && \leqq 2^{\nu-2} \\
&\quad\vdots && \quad\vdots && \\
&\text{rank } (\lambda_1 \lambda_2 \cdots \lambda_r) && \leqq \tfrac{1}{2} \text{ rank } (\lambda_1 \lambda_2 \cdots \lambda_{r-1}) && \leqq 2^{\nu-r} \\
&\quad\vdots && \quad\vdots && \\
&\text{rank } (\lambda_1 \lambda_2 \cdots \lambda_\nu) && \leqq \tfrac{1}{2} \text{ rank } (\lambda_1 \lambda_2 \cdots \lambda_{\nu-1}) && \leqq 2^{\nu-\nu} = 1.
\end{aligned} \tag{9.4}$$

However, if $A^{(i)}$ is a coördinate tensor of $A = L_1 + L_2 + \cdots + L_\nu$, then $\dfrac{1}{|i|!} A^{(i)} s_{(i)} = \lambda_1 \lambda_2 \cdots \lambda_\nu \neq 0$ and consequently rank $(\lambda_1 \lambda_2 \cdots \lambda_\nu) = 1$. Hence the inequality sign cannot hold at any step and rank $(\lambda_1 \lambda_2 \cdots \lambda_{|i|}) = 2^{\nu - |i|}$.

[9] C. C. MacDuffee, *The Theory of Matrices*, Berlin 1933, Theorem 8.3.

THEOREM (9.2). *If $A^{(i)}(|i| = \nu)$ is the coördinate tensor of an axis on the quadric and* $\alpha = \frac{1}{|i|!} A^{(i)} s_{(i)}$, *then*

$$\alpha \equiv \| \alpha^A{}_B \| = \| \psi^A \varphi_B \|. \tag{9.5}$$

By Theorem (9.1) the rank of α is one and hence all the columns of the matrix are proportional to any non-vanishing column with elements $\psi^1, \psi^2, \cdots, \psi^{2^\nu}$. If the factors of proportionality are called φ_B, we have (9.5). Interpreting ψ^A as the coördinates of a point in the spin space we have a representation

$$A \to \psi \tag{9.6}$$

of the axes on the quadric in $P_{2\nu-1}$ by means of points in $P_{2^\nu-1}$. In the following sections we shall study the figures in the spin space determined by linear spaces on the quadric and apply these results to the representation (9.6).

10. Spaces in $P_{2^\nu-1}$ Determined by Spaces on the Quadric in $P_{2\nu-1}$. The matrix $\beta = \frac{1}{|i|!} B^{(i)} s_{(i)}$ corresponding to a linear space B with coördinate tensor $B^{(i)}$ is, by Theorem (9.1), singular if B lies on the quadric. Hence the collineation determined by β transforms all the points of $P_{2^\nu-1}$ into the points of a subspace R_β, the "rank" space of β. A point ψ with coördinates ψ^A will be contained in R_β if and only if there exists a coördinate vector θ^B such that $\psi^A = \beta^A{}_B \theta^B$.

The singular points of the collineation form a linear space N_β, the "null" space of β. Thus a point ψ belongs to N_β if and only if $\beta\psi = 0$. It is then evident that $D(R_\beta) = \text{rank}\,\beta$ and $D(N_\beta) = \text{nullity of}\,\beta$, so that

$$D(R_\beta) + D(N_\beta) = 2^\nu. \tag{10.1}$$

In the following theorems of this and the next two sections, we shall suppose that B is a linear space on the quadric and that β is the matrix corresponding to B under (8.9).

THEOREM (10.1). *R_β is a subspace of N_β.*

If ψ belongs to R_β, $\psi = \beta\theta$. But $\beta^2 = 0$, so that $\beta\psi = \beta^2\theta = 0$ and hence ψ also belongs to N_β. The space R_β will be a proper subspace of N_β if the rank of β is less than the nullity of β and the spaces will coincide if the rank and nullity are equal. By Theorem (9.1) the latter case will arise only when B is a point and then R_β and N_β are coincident axes. Since Trace $(X^i\gamma_i) = 0$, the pointwise invariant spaces of a non-singular involution of the family $X^i\gamma_i$ are of the same dimensionality. They are therefore a pair of non-intersecting axes in $P_{2^\nu-1}$. The spaces R_β and N_β may be regarded as the coincident axes of the singular involution corresponding to a point on Q.

In a cartesian coördinate system the matrix

$$\gamma_0 = (-1)^{\frac{1}{2}\nu}\gamma_1\gamma_2\gamma_3 \cdots \gamma_{2\nu} \tag{10.2}$$

satisfies the equations

$$(10.3) \qquad (\gamma_0)^2 = 1 \quad \text{and} \quad \gamma_0\gamma_i = -\gamma_i\gamma_0.$$

Expressing γ_0 in the form (8.4), we could easily show that these equations determine γ_0 to within sign. It is evident that γ_0 continues to satisfy (10.3) if γ_i is replaced by $t_i^j\gamma_j$, where $|t_i^j| \neq 0$, and hence the involution defined by the pencil of matrices $\rho\gamma_0$ is uniquely determined by the linear family $X^i\gamma_i$. We shall call it the "invariant involution."

The invariant nature of the matrix γ_0 is put into evidence if we define

$$(10.4) \qquad \gamma_0 = \frac{(-1)^{\frac{1}{2}\nu} \mid \gamma_{ij} \mid^{-\frac{1}{2}}}{(2\nu)!} \epsilon^{(i)} \gamma_{i_1} \gamma_{i_2} \cdots \gamma_{i_{2\nu}},$$

which we may do since the right member reduces in a cartesian coördinate system to (10.2). The advantage of (10.4) is that it determines a particular one of the two matrices satisfying (10.3)—at least to within a choice of one of the square roots of the determinant $|\gamma_{ij}|$.

If this root is the same as that used in the definition of $X_{(\bar{i})}$ in terms of X^j (cf. (5.4)), γ_0 satisfies the equations

$$(10.5) \qquad \gamma_0 s_{(i)} = (-1)^{\frac{1}{2}(\nu+|i|)} s_{(\bar{i})}$$

and

$$(10.6) \qquad \gamma_0 \left(\frac{1}{|i|!} X^{(i)} s_{(i)} \right) = (-1)^{\frac{1}{2}(\nu-|i|)} \left(\frac{1}{|j|!} X^{(j)}_{-} s_{(j)} \right).$$

These equations are most easily verified in a cartesian coördinate system and their tensor form then insures their validity in all projective coördinate systems in $P_{2\nu-1}$. The corresponding formulas for multiplication on the right by γ_0 are obtained from these by using the invariant relations

$$(10.7) \qquad \gamma_0 s_{(i)} = (-1)^{|i|} s_{(i)}\gamma_0.$$

THEOREM (10.2). *The invariant involution leaves R_β and N_β each invariant.*

Since $\beta = \frac{1}{|i|!} B^{(i)} s_{(i)}$, where $B^{(i)}$ is a coördinate tensor of the space B on the quadric, we can use (10.7) to get $\gamma_0\beta = (-1)^{|i|}\beta\gamma_0$. If ψ is a point of R_β so that $\psi^B = \beta^B{}_C\theta^C$, then $\gamma^A{}_{0B}\psi^B = \beta^A{}_B[(-1)^{|i|}\gamma^B_{0C}\psi^C]$ and hence the involution transforms ψ into a point which again belongs to R_β. The invariance of N_β follows similarly.

11. The Invariant Polarity. Using the numerical values of the matrices γ_i which we gave in (7.4) and (7.5), it is easily proved that the matrix

$$(11.1) \qquad C = \gamma_1\gamma_3\gamma_5 \cdots \gamma_{2\nu-1}$$

satisfies the equations

$$(11.2) \qquad C' = eC, \qquad e = (-1)^{\frac{1}{2}\nu(\nu+1)}$$

and

$$(11.3)\qquad (C\gamma_0)' = fC\gamma_0, \quad (C\gamma_i)' = f(C\gamma_i), \qquad f = (-1)^\nu e.$$

These equations will remain valid in all projective coördinate systems in $P_{2^\nu-1}$ if we take C to be a matrix of the type[10] $C = \| C_{AB} \|$. We assume that C behaves as a scalar under transformations of coördinates in $P_{2^\nu-1}$ so that equations (11.3) remain valid after a transformation affecting the index i.

The equations

$$(11.4)\qquad \varphi_A = C_{AB}\psi^B$$

define a point $\to$ hyperplane transformation of the spin space which is a polarity in a quadric or in a linear complex, depending on the value of ν. Equations (11.3) imply that this polarity is geometrically commutative with the invariant involution and with every involution of the linear family $X^i\gamma_i$. We shall call it the "invariant polarity."

THEOREM (11.1). *The invariant polarity interchanges R_β and N_β.*

Let us put $D(B) = r$ and take $B_1^i, B_2^i, \cdots, B_r^i$ to be coördinate vectors of r linearly independent points in B. Multiplying the second of equations (11.3) through by B_p^i $(p = 1, 2, \cdots, r)$ and putting $\beta_p = B_p^i\gamma_i$, we get $\beta_p'C' = fC\beta_p$, or, using (11.2)

$$(11.5)\qquad C\beta_p = \pm\, \beta_p'C.$$

Since $\| C_{AB} \|$ is non-singular, D (polar of R_β) $= 2^\nu - D(R_\beta)$ and combining this equation with (10.1) gives D (polar of R_β) $= D(N_\beta)$. Hence to show that the invariant polarity transforms R_β into N_β it will be sufficient to prove that the polar of R_β contains N_β. This will be the case if the polar hyperplane of an arbitrary point ψ of R_β contains every point of N_β. We shall therefore have the theorem if we can prove that $\varphi^A C_{AB}\psi^B = 0$ whenever $\psi^B = \beta^B{}_C\theta^C$ and $\beta^D{}_A\varphi^A = 0$.

Calling $\| \psi^A \|$, $\| \varphi^A \|$ and $\| \theta^A \|$ matrices of one column, the condition is that $\varphi'C\psi = 0$ whenever $\psi = \beta\theta$ and $\beta\varphi = 0$. But since $\beta = \beta_1\beta_2 \cdots \beta_r$ we can employ (11.5) to get

$$\varphi'C\psi = \varphi'C\beta\theta = \varphi'C\beta_1\beta_2 \cdots \beta_r\theta = \pm\, (\varphi'\beta_1'\beta_2' \cdots \beta_r')C\theta.$$

The points B_p are conjugate by pairs in the quadric and consequently the matrices β_p anticommute by pairs. Hence

$$(\varphi'\beta_1'\beta_2' \cdots \beta_r') = (\beta_r \cdots \beta_2\beta_1\varphi)' = \pm\, (\beta_1\beta_2 \cdots \beta_r\varphi)' = \pm\, (\beta\varphi)' = 0$$

and substituting in the last equation gives the required condition, $\varphi'C\psi = 0$.

12. Properties of R_β and N_β. The matrix γ_0 satisfies the equation $x^2 - 1 = 0$, and consequently its roots are $+1$ and -1 and its elementary divisors are

---

[10] It can be shown that equations (11.3) possess a solution only for the value of f given and that they determine the matrix C to within a factor.

simple. Since trace $\gamma_0 = 0$, the sum of the roots is zero. A suitable choice of coördinate system in the spin space will therefore enable us to take γ_0 in the form

$$\gamma_0 = \left\| \begin{matrix} 1_{2^{\nu-1}} & 0 \\ 0 & -1_{2^{\nu-1}} \end{matrix} \right\|. \tag{12.1}$$

The pointwise invariant spaces of the involution defined by γ_0 are spaces of $2^{\nu-1} - 1$ dimensions which we denote by $[\gamma_0]^+$ and $[\gamma_0]^-$. The points of these axes are characterized by the equations

$$[\gamma_0]^+ : \gamma_{0\,B}^{\,A}\psi^B = +\psi^A,$$

(12.2) and

$$[\gamma_0]^- : \gamma_{0\,B}^{\,A}\psi^B = -\psi^A,$$

respectively. The representation (9.6) images axes on Q by points of $[\gamma_0]^+$ and $[\gamma_0]^-$ in the way stated in the

THEOREM (12.1). *The equations*

$$\frac{1}{\nu!} A^{(i)} s_{(i)B}^{\,A} = \psi^A C_{BC} \psi^C \tag{12.3}$$

and their inverses

$$A^{(i)} = (-1)^{\nu} C_{AC} s^{(i)C}_{\quad B} \psi^A \psi^B, \tag{12.4}$$

establish a (1-1) *correspondence*

$$A \leftrightarrow \psi \tag{12.5}$$

between the axes on the quadric in $P_{2\nu-1}$ and points in $P_{2^\nu-1}$. Under this correspondence axes belonging to F_+ correspond to points of the axis $[\gamma_0]^+$ and axes belonging to F_- correspond to points belonging to the axis $[\gamma_0]^-$.

If A is an axis on the quadric in $P_{2\nu-1}$, its corresponding matrix is, by Theorem (9.2), of the form $\alpha = \| \psi^A \varphi_B \|$ where R_β is the point with coördinates ψ^A and N_β is the hyperplane with coördinates φ_B. By Theorem (11.1) the vector φ_B is determined as a multiple of $C_{BC}\psi^C$ and therefore the point ψ determines the matrix α to within a factor. But the matrix α determines a coördinate tensor of A by the equations (8.8) and hence if ψ corresponds to any axis under (9.6) it uniquely determines the axis.

To prove the last part of the theorem we apply (10.6) to the matrix $\alpha = \frac{1}{|i|!} A^{(i)} s_{(i)}$ to get $\gamma_0 \alpha = \pm \alpha$ where the plus sign is used if $A^{(i)} = +A^{(i)}_-$ and the minus sign if $A^{(i)} = -A^{(i)}_-$. Hence $\gamma_{0\,B}^{\,A}\psi^B = +\psi^A$ if A belongs to F_+, and $\gamma_{0\,B}^{\,A}\psi^B = -\psi^A$ if A belongs to F_-. Theorem (10.2) could have been used to prove that ψ lies in either $[\gamma_0]^+$ or $[\gamma_0]^-$, but the argument just given proves in addition that points corresponding to axes of the same family lie in the same axis of γ_0 and points corresponding to axes of different families lie in different axes of γ_0.

The invariance of R_β under γ_0 implies that R_β intersects $[\gamma_0]^+$ and $[\gamma_0]^-$ in spaces $R_\beta[\gamma_0]^+$ and $R_\beta[\gamma_0]^-$ such that

$$(12.6) \qquad D(R_\beta[\gamma_0]^+) + D(R_\beta[\gamma_0]^-) = D(R_\beta).$$

If B is not an axis, the statement that R_β intersects $[\gamma_0]^+$ and $[\gamma_0]^-$ in spaces of the same dimensionality is contained in the

THEOREM (12.2). *If $D(R_\beta) > 1$, then*

$$(12.7) \qquad D(R_\beta[\gamma_0]^+) = D(R_\beta[\gamma_0]^-) = \tfrac{1}{2}D(R_\beta).$$

If B is not an axis, its polar space does not lie entirely on the quadric and hence we may choose a point in it in such a way that its corresponding collineation Γ is non-singular. Of course Γ commutes (or anticommutes) with the collineation β corresponding to B. If $\psi = \beta\theta$ is any point of R_β, $\Gamma\psi = \Gamma\beta\theta = \pm\beta(\Gamma\theta)$ and hence Γ leaves R_β invariant. Since Γ is an involution of the family $X^i\gamma_i$, it anticommutes with γ_0 and hence interchanges $[\gamma_0]^+$ and $[\gamma_0]^-$. Therefore Γ interchanges $R_\beta[\gamma_0]^+$ and $R_\beta[\gamma_0]^-$ and the spaces are of the same dimensionality.

An $(r + 1)$-dimensional linear space B_1 on the quadric is determined as the join of an r-dimensional subspace B and a point L_1 in B_1 but not in B. The spaces corresponding to B and L_1 determine the spaces corresponding to B_1 in the way stated in the

THEOREM (12.3). *If $B_1 = B + L_1$ is a linear space on the quadric, L_1 is a point, $BL_1 = 0$, and B_1, B and L_1 correspond to matrices β_1, β and λ_1 respectively, then*

$$(12.8) \qquad R_{\beta_1} = R_\beta R_{\lambda_1} \quad \textit{and} \quad N_{\beta_1} = N_\beta + N_{\lambda_1}.$$

The second of these equations is obtained from the first by taking polars with respect to the invariant polarity. Moreover, since $\beta_1 = \beta\lambda_1 = \pm\lambda_1\beta$ it is evident that R_{β_1} is included in both R_β and R_{λ_1} and therefore in $R_\beta R_{\lambda_1}$. Hence it is sufficient to prove that $D(R_\beta R_{\lambda_1}) = D(R_{\beta_1})$.

Let us choose L_2 to be a point on the intersection of the quadric Q with the polar of B. This can always be done in such a way that L_2 is not conjugate to L_1 and then $B_2 = B + L_2$ will lie on Q and the line $L_1 + L_2$ will not lie on Q. We now take M_1 and M_2 to be two distinct points on $L_1 + L_2$ which are conjugate with respect to Q and call their corresponding matrices μ_1 and μ_2. Then $\mu_1 = a\lambda_1 + b\lambda_2$ and $\mu_2 = c\lambda_1 + d\lambda_2$. Since both λ_1 and λ_2 commute with β so do μ_1 and μ_2, and hence the non-singular involutions with matrices μ_1 and μ_2 transform R_β into itself. The same is therefore true of their product, $\mu_1\mu_2$. Making use of the relation $\mu_1\mu_2 = -\mu_2\mu_1$, a computation shows that the pointwise invariant spaces of the involution determined by $\mu_1\mu_2$ are the rank spaces of the singular elements, λ_1 and λ_2, of the pencil $\rho\lambda_1 + \sigma\lambda_2$. Hence $D(R_\beta R_{\lambda_1}) + D(R_\beta R_{\lambda_2}) = D(R_\beta)$. The involution with matrix μ_1 interchanges R_{λ_1} and R_{λ_2} and leaves R_β invariant so that $D(R_\beta R_{\lambda_1}) = D(R_\beta R_{\lambda_2})$. Therefore $D(R_\beta R_{\lambda_1}) = \frac{1}{2}D(R_\beta)$, which is equal to $D(R_{\beta_1})$ by Theorem (9.1).

13. Geometry of a Generalization of the Plücker-Klein Correspondence. The properties of the geometrical correspondence

$$B \to R_\beta \tag{13.1}$$

between linear spaces on Q and the spaces into which the corresponding matrices transform the whole of $P_{2^\nu-1}$ will be made clearer if we discuss some special cases.

When $\nu = 1$ the quadric in P_1 consists of two distinct points and these points correspond, respectively, to two distinct points in the spin space, which is again a P_1.

When $\nu = 2$ the spaces $P_{2\nu-1}$ and $P_{2^\nu-1}$ are again of the same number of dimensions but the axes on the quadric are now represented by points of two non-intersecting lines l_1 and l_2 in the spin space P_3. Two lines on the quadric which intersect in a point P_0 correspond to two points L_1 and L_2 lying on the lines l_1 and l_2 respectively, and the point P_0 corresponds to the line $L_1 + L_2$ crossing l_1 and l_2.

When $\nu = 3$ the quadric is in P_5 and the spin space is 7-dimensional. The axes $[\gamma_0]^+$ and $[\gamma_0]^-$ are non-intersecting 3-spaces. A point on Q corresponds to a P_3 in P_7 which intersects each of the 3-spaces $[\gamma_0]^+$ and $[\gamma_0]^-$ in a line. Two points B_1 and B_2 on Q therefore determine a pair of lines in $[\gamma_0]^+$ and a pair in $[\gamma_0]^-$. These two pairs of lines may intersect in points L_1 and L_2 and if they do the line $B_1 + B_2$ lies on Q and corresponds to the line $L_1 + L_2$ crossing $[\gamma_0]^+$ and $[\gamma_0]^-$. If the plane $A = B_1 + B_2 + B_3$ determined by the three points B_1, B_2 and B_3 lies on Q and belongs to F_+, then the points determine in $[\gamma_0]^-$ the edges of a non-degenerate triangle while in $[\gamma_0]^+$ they determine three lines through the point corresponding to A under (13.1).

For $\nu = 1, 2$, or 3 it can be shown that all the points of $[\gamma_0]^+$ and $[\gamma_0]^-$ occur as images of axes on Q. When $\nu > 3$ this is no longer the case, as follows from the fact that the dimension of $[\gamma_0]^+$ is $2^{\nu-1} - 1$ while the dimension[11] of the family of axes F^+ is $\frac{1}{2}\,\nu(\nu - 1)$.

In a coördinate system in which γ_0 is given by (12.1), equations (11.2) and (11.3) give $\gamma_0 C = (-1)^\nu C\gamma_0$ and hence C is of the form

$$C = \begin{Vmatrix} C_1 & 0 \\ 0 & C_2 \end{Vmatrix}, \qquad C_1' = e\,C_1, \qquad C_2' = e\,C_2, \tag{13.2}$$

when ν is even, and of the form

$$C = \begin{Vmatrix} 0 & C_1 \\ C_2 & 0 \end{Vmatrix}, \qquad C_2' = e\,C_1 \tag{13.3}$$

when ν is odd. The invariant polarity therefore interchanges the axes $[\gamma_0]^+$ and $[\gamma_0]^-$ if ν is even and leaves them separately invariant if ν is odd.

When ν is even a point of $[\gamma_0]^+$ is transformed into a hyperplane which contains

[11] Bertini, *loc. cit.*, p. 142.

The invariance of R_β under γ_0 implies that R_β intersects $[\gamma_0]^+$ and $[\gamma_0]^-$ in spaces $R_\beta[\gamma_0]^+$ and $R_\beta[\gamma_0]^-$ such that

$$(12.6)\qquad D(R_\beta[\gamma_0]^+) + D(R_\beta[\gamma_0]^-) = D(R_\beta).$$

If B is not an axis, the statement that R_β intersects $[\gamma_0]^+$ and $[\gamma_0]^-$ in spaces of the same dimensionality is contained in the

THEOREM (12.2). *If $D(R_\beta) > 1$, then*

$$(12.7)\qquad D(R_\beta[\gamma_0]^+) = D(R_\beta[\gamma_0]^-) = \tfrac{1}{2}D(R_\beta).$$

If B is not an axis, its polar space does not lie entirely on the quadric and hence we may choose a point in it in such a way that its corresponding collineation Γ is non-singular. Of course Γ commutes (or anticommutes) with the collineation β corresponding to B. If $\psi = \beta\theta$ is any point of R_β, $\Gamma\psi = \Gamma\beta\theta = \pm\beta(\Gamma\theta)$ and hence Γ leaves R_β invariant. Since Γ is an involution of the family $X^i\gamma_i$, it anticommutes with γ_0 and hence interchanges $[\gamma_0]^+$ and $[\gamma_0]^-$. Therefore Γ interchanges $R_\beta[\gamma_0]^+$ and $R_\beta[\gamma_0]^-$ and the spaces are of the same dimensionality.

An $(r + 1)$-dimensional linear space B_1 on the quadric is determined as the join of an r-dimensional subspace B and a point L_1 in B_1 but not in B. The spaces corresponding to B and L_1 determine the spaces corresponding to B_1 in the way stated in the

THEOREM (12.3). *If $B_1 = B + L_1$ is a linear space on the quadric, L_1 is a point, $BL_1 = 0$, and B_1, B and L_1 correspond to matrices β_1, β and λ_1 respectively, then*

$$(12.8)\qquad R_{\beta_1} = R_\beta R_{\lambda_1} \quad \textit{and} \quad N_{\beta_1} = N_\beta + N_{\lambda_1}.$$

The second of these equations is obtained from the first by taking polars with respect to the invariant polarity. Moreover, since $\beta_1 = \beta\lambda_1 = \pm\lambda_1\beta$ it is evident that R_{β_1} is included in both R_β and R_{λ_1} and therefore in $R_\beta R_{\lambda_1}$. Hence it is sufficient to prove that $D(R_\beta R_{\lambda_1}) = D(R_{\beta_1})$.

Let us choose L_2 to be a point on the intersection of the quadric Q with the polar of B. This can always be done in such a way that L_2 is not conjugate to L_1 and then $B_2 = B + L_2$ will lie on Q and the line $L_1 + L_2$ will not lie on Q. We now take M_1 and M_2 to be two distinct points on $L_1 + L_2$ which are conjugate with respect to Q and call their corresponding matrices μ_1 and μ_2. Then $\mu_1 = a\lambda_1 + b\lambda_2$ and $\mu_2 = c\lambda_1 + d\lambda_2$. Since both λ_1 and λ_2 commute with β so do μ_1 and μ_2, and hence the non-singular involutions with matrices μ_1 and μ_2 transform R_β into itself. The same is therefore true of their product, $\mu_1\mu_2$. Making use of the relation $\mu_1\mu_2 = -\mu_2\mu_1$, a computation shows that the pointwise invariant spaces of the involution determined by $\mu_1\mu_2$ are the rank spaces of the singular elements, λ_1 and λ_2, of the pencil $\rho\lambda_1 + \sigma\lambda_2$. Hence $D(R_\beta R_{\lambda_1}) + D(R_\beta R_{\lambda_2}) = D(R_\beta)$. The involution with matrix μ_1 interchanges R_{λ_1} and R_{λ_2} and leaves R_β invariant so that $D(R_\beta R_{\lambda_1}) = D(R_\beta R_{\lambda_2})$. Therefore $D(R_\beta R_{\lambda_1}) = \frac{1}{2}D(R_\beta)$, which is equal to $D(R_{\beta_1})$ by Theorem (9.1).

13. **Geometry of a Generalization of the Plücker-Klein Correspondence.** The properties of the geometrical correspondence

$$B \to R_\beta \tag{13.1}$$

between linear spaces on Q and the spaces into which the corresponding matrices transform the whole of $P_{2^\nu-1}$ will be made clearer if we discuss some special cases.

When $\nu = 1$ the quadric in P_1 consists of two distinct points and these points correspond, respectively, to two distinct points in the spin space, which is again a P_1.

When $\nu = 2$ the spaces $P_{2\nu-1}$ and $P_{2^\nu-1}$ are again of the same number of dimensions but the axes on the quadric are now represented by points of two non-intersecting lines l_1 and l_2 in the spin space P_3. Two lines on the quadric which intersect in a point P_0 correspond to two points L_1 and L_2 lying on the lines l_1 and l_2 respectively, and the point P_0 corresponds to the line $L_1 + L_2$ crossing l_1 and l_2.

When $\nu = 3$ the quadric is in P_5 and the spin space is 7-dimensional. The axes $[\gamma_0]^+$ and $[\gamma_0]^-$ are non-intersecting 3-spaces. A point on Q corresponds to a P_3 in P_7 which intersects each of the 3-spaces $[\gamma_0]^+$ and $[\gamma_0]^-$ in a line. Two points B_1 and B_2 on Q therefore determine a pair of lines in $[\gamma_0]^+$ and a pair in $[\gamma_0]^-$. These two pairs of lines may intersect in points L_1 and L_2 and if they do the line $B_1 + B_2$ lies on Q and corresponds to the line $L_1 + L_2$ crossing $[\gamma_0]^+$ and $[\gamma_0]^-$. If the plane $A = B_1 + B_2 + B_3$ determined by the three points B_1, B_2 and B_3 lies on Q and belongs to F_+, then the points determine in $[\gamma_0]^-$ the edges of a non-degenerate triangle while in $[\gamma_0]^+$ they determine three lines through the point corresponding to A under (13.1).

For $\nu = 1, 2,$ or 3 it can be shown that all the points of $[\gamma_0]^+$ and $[\gamma_0]^-$ occur as images of axes on Q. When $\nu > 3$ this is no longer the case, as follows from the fact that the dimension of $[\gamma_0]^+$ is $2^{\nu-1} - 1$ while the dimension[11] of the family of axes F^+ is $\frac{1}{2}\,\nu(\nu - 1)$.

In a coördinate system in which γ_0 is given by (12.1), equations (11.2) and (11.3) give $\gamma_0 C = (-1)^\nu C\gamma_0$ and hence C is of the form

$$C = \begin{Vmatrix} C_1 & 0 \\ 0 & C_2 \end{Vmatrix}, \qquad C_1' = e\,C_1, \qquad C_2' = e\,C_2, \tag{13.2}$$

when ν is even, and of the form

$$C = \begin{Vmatrix} 0 & C_1 \\ C_2 & 0 \end{Vmatrix}, \qquad C_2' = e\,C_1 \tag{13.3}$$

when ν is odd. The invariant polarity therefore interchanges the axes $[\gamma_0]^+$ and $[\gamma_0]^-$ if ν is even and leaves them separately invariant if ν is odd.

When ν is even a point of $[\gamma_0]^+$ is transformed into a hyperplane which contains

---

[11] Bertini, *loc. cit.*, p. 142.

$[\gamma_0]^-$ and intersects $[\gamma_0]^+$ in a space of $2^{\nu-1} - 2$ dimensions. The polar transformation of $P_{2^\nu-1}$ induces a polarity within the space $[\gamma_0]^+$ and, similarly, within $[\gamma_0]^-$.

When ν is odd, a point of $[\gamma_0]^+$ is transformed into a hyperplane which contains $[\gamma_0]^-$ and intersects $[\gamma_0]^+$ in a space of $2^{\nu-1} - 2$ dimensions. The invariant polarity therefore defines a mapping of the subspaces of $[\gamma_0]^+$ into subspaces of $[\gamma_0]^-$ in which points correspond to spaces of $2^{\nu-1} - 2$ dimensions, lines correspond to spaces of $2^{\nu-1} - 3$ dimensions, and so on. In particular, if R_α is an axis in $P_{2^\nu-1}$ corresponding to a point on Q, the mapping carries $R_\alpha[\gamma_0]^+$ into $R_\alpha[\gamma_0]^-$. Since $R_\alpha = R_\alpha[\gamma_0]^+ + R_\alpha[\gamma_0]^-$, the space R_α is determined by its intersection with $[\gamma_0]^+$. Similarly, a point of $[\gamma_0]^-$ is determined by the $(2^{\nu-1} - 2)$-dimensional space in which its polar hyperplane intersects $[\gamma_0]^+$. These results can be combined with the theorems of §12 to give the

THEOREM (13.1). *If ν is odd and $\geqq 3$, the points of a quadric Q in $P_{2\nu-1}$ may be made to correspond to axes in $P_{2^\nu-1} - 1$ in such a way that the points of an $(r - 1)$-dimensional space, $r < \nu$, on Q correspond to axes all of which contain the same $(2^{\nu-1-r} - 1)$-dimensional space. Under the correspondence the points of an axis on Q either correspond to axes all of which contain the same point or to axes all of which lie in the same hyperplane.*

For $\nu = 3$ this theorem gives the Plücker-Klein correspondence between points on a quadric in P_5 and lines in P_3. In this case all the lines in P_3 enter into the correspondence. Lines on the quadric correspond to pencils of lines in P_3 and the points of a plane on the quadric correspond to all the lines through a point or to all the lines in a plane.

14. Collineation Representation of $H_{2\nu}^+$ for $\nu > 2$.

The geometrical correspondences established in this chapter give rise to representations of the proper orthogonal group in 2ν variables, $H_{2\nu}^+$. In this and the next two sections we shall discuss these representations.

A collineation of the spin space, with matrix $P = \|\, P^A{}_B \,\|$, will leave invariant the linear family of involutions $X^i\gamma_i$ if

$$P(X^s\gamma_s)P^{-1} = Y^r\gamma_r \tag{14.1}$$

for arbitrary values of the parameters X^i of the family. From (9.2), trace $\gamma^i\gamma_j = 2^\nu\delta^i_j$. Hence, multiplying both members of (14.1) by γ^i and taking the trace gives

$$Y^i = L^i_jX^j, \tag{14.2}$$

where

$$L^i_j = 2^{-\nu}\,\mathrm{Trace}\,(\gamma^iP\gamma_jP^{-1}). \tag{14.3}$$

The collineation with matrix P therefore transforms the matrices γ_i into

$$\eta_i \equiv P\gamma_iP^{-1} = \gamma_sL^s_i, \tag{14.4}$$

and induces in $P_{2^\nu-1}$ the collineation (14.2).

The quadric is left invariant by (14.2) since it is the locus of singular involutions of the linear family. Indeed, in a cartesian coördinate system we have

$$\sum (X^i)^2 1 = (X^i\gamma_i)^2 = P^{-1}(X^i\eta_i)^2 P = P^{-1}(Y^s\gamma_s)^2 P = \sum (Y^i)^2 1, \tag{14.5}$$

and consequently the linear transformation (14.2) leaves the quadratic form $\gamma_{ij}X^iX^j$ invariant as well as the quadric $\gamma_{ij}X^iX^j = 0$. Hence

$$\gamma_{rs}L_i^rL_j^s = \gamma_{ij}, \tag{14.6}$$

and taking determinants

$$| L_j^i | = \pm 1. \tag{14.7}$$

A matrix $\| L_j^i \|$ satisfying (14.6) will be called proper orthogonal if its determinant is $+ 1$ and improper orthogonal if its determinant is $- 1$. The set of all orthogonal matrices form the orthogonal group $H_{2\nu}$ and the proper orthogonal matrices constitute an invariant subgroup $H_{2\nu}^+$.

If $\| L_j^i \|$ satisfies (14.6) and the coördinate system is cartesian so that γ_i satisfy (7.3), then the matrices $\eta_i = \gamma_s L_i^s$ also satisfy (7.3). A successive specialization of the coördinate system will suffice to show that any ordered set of 2ν matrices of order 2^ν satisfying (7.3) can be simultaneously transformed by a similarity into the canonical forms given by (7.4) and (7.5). Hence if $\| L_j^i \|$ is any orthogonal matrix there exists a matrix[12] $P = \| P^A{}_B \|$ such that $\gamma_s L_i^s = P\gamma_i P^{-1}$.

Moreover, if $R\gamma_i R^{-1} = P\gamma_i P^{-1}$, then $P^{-1}R$ commutes with the matrices γ_i and consequently with all the matrices (8.3). Since these form a basis for the full matrix algebra, $P^{-1}R = \rho 1$ and $R = \rho P$. The orthogonal matrix $\| L_j^i \|$ therefore determines P to within a factor by means of equations (14.4). This completes the proof of the

THEOREM (14.1). *Equations (14.4) (or (14.3)) establish a (1-1) isomorphism*

$$\| L_j^i \| \leftrightarrow \rho \, \| P^A{}_B \| \tag{14.8}$$

between the group of orthogonal matrices of order 2ν and the group of collineations of $P_{2^\nu-1}$ which leave the linear family $X^i\gamma_i$ invariant.

Using (14.4) and the definition of γ_0, a computation gives

$$P\gamma_0 P^{-1} = | L_j^i | \gamma_0 \tag{14.9}$$

and hence if P corresponds to a proper orthogonal matrix it commutes with γ_0 while if it corresponds to an improper orthogonal matrix it anticommutes with γ_0. In the coördinate system of (12.1),

$$P = \begin{Vmatrix} P_+ & 0 \\ 0 & P_- \end{Vmatrix} \tag{14.10}$$

[12] This also follows from the theorem that the full matrix algebra allows only inner automorphisms. For the details of this proof see the paper of Brauer and Weyl cited above.

if the collineation corresponds to a transformation of $H_{2\nu}^+$, and

$$P = \begin{Vmatrix} 0 & P_1 \\ P_2 & 0 \end{Vmatrix} \tag{14.11}$$

if the collineation corresponds to an improper transformation of $H_{2\nu}$.

It is evident that collineations of the type (14.10) leave $[\gamma_0]^+$ and $[\gamma_0]^-$ separately invariant while collineations of the form (14.11) interchange them. Using the (1-1) correspondence of Theorem (12.1) between axes on the quadric and points in $[\gamma_0]^+$ and $[\gamma_0]^-$ we get the

THEOREM (14.2). *Collineations in $P_{2\nu-1}$ defined by proper orthogonal matrices ($|L_j^i| = +1$) leave the two families of axes on the quadric separately invariant while collineations defined by improper orthogonal matrices interchange them.*

If all the matrices of a group are of the form (14.10), elementary arguments suffice to show that the matrices P_+ form a group which is isomorphic (perhaps multiply) with the original group. Combining this isomorphism with (14.8) we get the isomorphism

$$\| L_j^i \| \to \rho P_+ \tag{14.12}$$

between $H_{2\nu}^+$ and a collineation group in $P_{2^{\nu-1}-1}$. Moreover, two different matrices $\| L_j^i \|$ and $\| M_j^i \|$ which correspond to the same pencil ρP_+, induce the same permutation of points in $[\gamma_0]^+$ and hence the transformations of $P_{2\nu-1}$ which they define effect the same permutation of the axes of F_+. If $\nu > 2$, Theorem (6.7), states that the two transformations must be identical. This will be the case only if $\| L_j^i \|$ and $\| M_j^i \|$ are proportional. Since both the matrices are orthogonal, $\| M_j^i \| = \pm \| L_j^i \|$.

The matrix γ_0 anticommutes with all the matrices γ_i so that $\gamma_0\gamma_i\gamma_0^{-1} = -\gamma_i$. Comparing this with (14.4) we see that under (14.8)

$$\| -\delta_j^i \| \leftrightarrow \rho\gamma_0 = \rho \begin{Vmatrix} 1 & 0 \\ 0 & -1 \end{Vmatrix}. \tag{14.13}$$

Hence if $\| L_j^i \|$ corresponds to ρP_+ under (7.8), $\| -L_j^i \|$ also corresponds to the same pencil of matrices. The isomorphism inverse to (7.8) is therefore (1-2) and we have the

THEOREM (14.3). *If $\nu > 2$, the matrix group $H_{2\nu}^+$ is (2-1) isomorphic with a collineation group in $P_{2^{\nu-1}-1}$.*

The theorem cannot be extended to the case $\nu = 2$ since the proper quadric group in three dimensional space is a six parameter group while the entire collineation group on the line involves only three parameters.[13] When $\nu = 3$

[13] See, however, *Geometry of Two-Component Spinors* by O. Veblen, (Proc. Nat. Acad., 19, 1933, pp. 462–474) where it is proved that the proper collineation group of a real non-ruled quadric in 3-space is (1-1) isomorphic with the collineation group on the complex projective line. The geometric correspondence underlying this isomorphism is the representation of the points of the complex projective line by means of the points of a sphere in real 3-space.

the theorem gives the group isomorphism associated with the Plücker-Klein representation of the lines of 3-space by points of a quadric in P_5.

15. **Matrix Representation of $H_{2\nu}^{+}$ for $\nu > 2$.** Equations (14.4) and (14.9) may be combined into the equations

$$P\gamma_\alpha P^{-1} = \gamma_\beta L_\alpha^\beta, \qquad (\alpha, \beta = 0, 1, \cdots, 2\nu) \tag{15.1}$$

if we define $L_i^0 = L_0^i = 0$ and $L_0^0 = 1$. Equations (14.3) combine to give

$$(C\gamma_\beta)' = f(C\gamma_\beta), \qquad (\beta = 0, 1, \cdots, 2\nu) \tag{15.2}$$

and multiplying through by L_α^β and summing on β gives

$$(CP\gamma_\alpha P^{-1})' = f\, C\, P\, \gamma_\alpha P^{-1}, \text{ or } ((P'CP)\gamma_\alpha)' = f(P'CP)\gamma_\alpha.$$

Hence

$$\begin{aligned} C^{-1}(P'CP)\gamma_\alpha &= fC^{-1}((P'CP)\gamma_\alpha)' = f((P'CP)\gamma_\alpha C'^{-1})' \\ &= ((P'\,CP)C^{-1}\gamma_\alpha')' = \gamma_\alpha C^{-1}(P'CP), \end{aligned}$$

where we have used the relation $C' = e\,C$.

The matric $C^{-1}(P'CP)$ therefore commutes with all the matrices γ_i and is consequently a multiple of the identity matrix. That is, $P'CP = \rho C$ and the collineation with matrix P transforms the invariant polarity into itself. We may therefore normalize the matrices P entering into the isomorphism (14.8) by the requirement that

$$P'CP = C. \tag{15.3}$$

This normalization selects out of each pencil ρP a pair of matrices which differ only in sign.

The collineation representation (14.8) now becomes the (1-2) matrix representation

$$\| L_j^i \| \leftrightarrow \pm P = \pm \begin{Vmatrix} P_+ & 0 \\ 0 & P_- \end{Vmatrix} \tag{15.4}$$

of proper orthogonal matrices of order 2ν by matrices of order 2^ν. We may again conclude that the matrices $\pm P_+$ form a group isomorphic (perhaps multiply) to the group of matrices $\pm P$, and so obtain the isomorphism

$$\| L_j^i \| \rightarrow \pm P_+. \tag{15.5}$$

We shall refer to this correspondence between proper orthogonal matrices of order 2ν and matrices of order $2^{\nu-1}$ as the representation Δ_+.

We shall now determine the nature of the correspondence inverse to (15.5) for values of ν greater than two. Since the isomorphism of Theorem (14.3) is (2-1) the orthogonal matrices corresponding to a given pair of spin matrices $\pm P_+$ are at most the two matrices $L = \| L_j^i \|$ and $-L$. Moreover, if $\pm P_+$

corresponds to both $-L$ and $+L$, then $\pm(P_+)(P_+^{-1}) = \pm 1_{2^{\nu-1}}$ corresponds to $-L(L^{-1}) = -\|\delta_j^i\|$. Conversely, if $-\|\delta_j^i\|$ corresponds to $\pm 1_{2^{\nu-1}}$, then both L and $-L$ correspond to the same pair of spin matrices $\pm P_+$.

By (14.13) the orthogonal matrix $-\|\delta_j^i\|$ corresponds to two matrices (of order 2^ν) out of the pencil $\rho\gamma_0$. To determine the values of ρ for which the matrix $\rho\gamma_0$ satisfies the normalizing condition, we make use of equations (11.2) and (11.3) to get

$$[(-1)^{\frac{1}{2}\nu}\gamma_0]'C[(-1)^{\frac{1}{2}\nu}\gamma_0] = C. \tag{15.6}$$

Hence we have for ν odd,

$$\|-\delta_j^i\| \to \pm i 1_{2^{\nu-1}}; \tag{15.7}$$

and for ν even,

$$\|-\delta_j^i\| \to \pm 1_{2^{\nu-1}}. \tag{15.8}$$

The properties of the representation Δ_+ therefore depend essentially upon whether ν is odd or even.

THEOREM (15.1). *If $\nu > 2$ is odd, Δ_+ is a (1-2) representation in which*

$$\|L_j^i\| \leftrightarrow \pm P_+,$$

(15.9) and

$$-\|L_j^i\| \leftrightarrow \pm iP_+.$$

The group $H_{2\nu}^+$ contains with each matrix $\|L_j^i\|$ its negative and no other matrices proportional to it; the group G of matrices of order $2^{\nu-1}$ contains with any matrix P_+ also the matrices $-P_+$, iP_+ and $-iP_+$ and no other matrices proportional to P_+. It is not possible to sharpen Δ_+ to a (1-1) representation. The group G does not contain a proper subgroup in which there occurs at least one out of each set of four matrices $\pm P_+$, $\pm iP_+$.

In order to sharpen Δ_+ to a (1-1) representation it would be necessary to select a single matrix of order $2^{\nu-1}$ to correspond to each of the proper orthogonal transformations

$$Y^1 = -X^1, \quad Y^2 = -X^2, \quad Y^3 = X^3, \quad Y^4 = X^4, \cdots, \quad Y^{2\nu} = X^{2\nu}, \tag{15.10}$$

and

$$Y^1 = -X^1, \quad Y^2 = X^2, \quad Y^3 = -X^3, \quad Y^4 = X^4, \cdots, \quad Y^{2\nu} = X^{2\nu}. \tag{15.11}$$

In a cartesian coördinate system these transformations correspond to collineations defined by $\rho\gamma_1\gamma_2$ and $\rho\gamma_1\gamma_3$. A (1-1) representation would therefore include two partial matrices, say $(a\gamma_1\gamma_2)_+$ and $(b\gamma_1\gamma_3)_+$, out of these pencils. These partial matrices anticommute while (15.10) and (15.11) commute, so that the (1-1) representation would break down for their product.

Any subgroup of G containing at least one out of every set of four proportional

matrices would contain the matrices $(a\gamma_1\gamma_2)_+$ and $(b\gamma_1\gamma_3)_+$ for some choice of the scalars a and b. Hence the subgroup would contain the matrix $(a\gamma_1\gamma_2)_+^{-1}$ $(b\gamma_1\gamma_3)_+^{-1}$ $(a\gamma_1\gamma_2)_+(b\gamma_1\gamma_3)_+$, which is equal to -1. If the subgroup contained three matrices out of any one pencil, it would contain all four matrices of the pencil and therefore all the elements of G.

The only remaining possibility is for the subgroup to contain one of the two pairs $\pm P_+$ and $\pm iP_+$ out of each set of four proportional matrices. Such a selection of matrices out of G would lead to the determination of a subgroup of $H_{2\nu}^+$ which would contain one out of every pair of orthogonal matrices $\pm \| L_j^i \|$. Such a subgroup cannot exist for it would necessarily contain either

$$\left\| \begin{array}{cc|cc|c|cc} 0 & 1 & & & & & \\ -1 & 0 & & & & & \\ \hline & & 0 & 1 & & & \\ & & -1 & 0 & & & \\ \hline & & & & \cdot & & \\ & & & & \cdot & & \\ \hline & & & & & 0 & 1 \\ & & & & & -1 & 0 \end{array} \right\| \tag{15.12}$$

or its negative and also the square of the matrix chosen, which is in both cases -1.

This proof shows, incidentally, that $H_{2\nu}^+$ is not the direct product of the two-element group consisting of 1 and -1 and a proper subgroup of itself. We may also conclude from Theorem (15.1) that no identification of the matrices $\| L_j^i \|$ and $- \| L_j^i \|$ can lead to a (2-2) isomorphism of the sort given in the following theorem.

THEOREM (15.2). *If $\nu > 2$ is even, Δ_+ is a (2-2) representation in which*

$$\pm \| L_j^i \| \leftrightarrow \pm P_+ . \tag{15.13}$$

It is not possible to sharpen Δ_+ to a (2-1), (1-2) or (1-1) representation.

The first part of the theorem follows immediately from (15.8) and the preceding discussion. A (2-1) representation cannot be obtained from Δ_+ since the group of matrices of order $2^{\nu-1}$ contains elements which anticommute. A (1-2) representation cannot exist since it would imply that $H_{2\nu}^+$ is the direct product of the two element group and a subgroup of itself. Finally, a (1-1) representation cannot be obtained from Δ_+ for the commutative transformations (15.10) and (15.11) correspond to anticommuting matrices.

16. **Matrix Representation of H_2^+ and H_4^+.** We complete the discussion of Δ_+ by treating the cases $\nu = 1$ and $\nu = 2$ in the two following theorems.

THEOREM (16.1). *If the quadratic form in E_2 is taken to be X^1X^2, the proper "orthogonal" matrices are of the form*

$$\| L^i_j \| = \begin{Vmatrix} a & 0 \\ 0 & 1/a \end{Vmatrix} \tag{16.1}$$

and the representation Δ_+ is the (1-2) correspondence

$$\| L^i_j \| \to \pm \sqrt{a}. \tag{16.2}$$

To prove this theorem we observe that when $\nu = 1$, $C = \begin{Vmatrix} 0 & c \\ -c & 0 \end{Vmatrix}$ so that (15.3) is the condition $| P | = +1$. Hence $P = \begin{Vmatrix} p & 0 \\ 0 & 1/p \end{Vmatrix}$. Taking $X^i\gamma_i = \begin{Vmatrix} 0 & X^1 \\ X^2 & 0 \end{Vmatrix}$, we have $(X^i\gamma_i)^2 = X^1X^2 1$. The equations of the isomorphism, $P(X^i\gamma_i)P^{-1} = Y^j\gamma_j$, are then

$$Y^1 = p^2X^1 \quad \text{and} \quad Y^2 = \frac{1}{p^2}X^2. \tag{16.3}$$

Hence $\quad \| L^i_j \| = \begin{Vmatrix} a & 0 \\ 0 & 1/a \end{Vmatrix} = \begin{Vmatrix} p^2 & 0 \\ 0 & 1/p^2 \end{Vmatrix}$ and $p = \pm \sqrt{a}$.

THEOREM (16.2). *A proper orthogonal matrix $\| L^i_j \|$ of order four corresponds under (15.4) to a pair of matrices*

$$\pm P = \pm \begin{Vmatrix} A & 0 \\ 0 & B \end{Vmatrix} \tag{16.4}$$

where $| A | = | B | = 1$. Conversely, if A and B are any two unimodular matrices of order two, then the matrices (16.4) correspond to a proper orthogonal matrix under (15.4). The representation Δ_+ is $(2\text{-}\infty^3)$.

When $\nu = 2$, the matrix C is skew-symmetric and since it is of the form (13.2) we must have

$$\left\| \begin{array}{cc|cc} 0 & c_1 & & \\ -c_1 & 0 & & \\ \hline & & 0 & c_2 \\ & & -c_2 & 0 \end{array} \right\| \tag{16.5}$$

so that the matrices (16.4) satisfy the normalizing condition if and only if $| A | = | B | = 1$.

It is a theorem of 3-dimensional projective geometry that a collineation may be constructed which simultaneously effects arbitrarily given collineations on the two reguli of lines on a quadric. Since we have represented the lines on Q by points of $[\gamma_0]^+$ and $[\gamma_0]^-$, the matrices A and B describe the collineations on the two reguli and may be taken to be any two unimodular matrices. We could also show this directly by proving that every matrix of the form (16.4) transforms the linear family of collineations into itself.

PRINCETON UNIVERSITY AND THE INSTITUTE FOR ADVANCED STUDY.

ANNALS OF MATHEMATICS
Vol. 38, No. 2, April, 1937

FUNCTIONAL TOPOLOGY AND ABSTRACT VARIATIONAL THEORY

BY MARSTON MORSE

(Received September 5, 1936)

INTRODUCTION

The critical points of a real single-valued differentiable function $f(x)$ of n real variables (x) are the points at which all of the partial derivatives of f vanish. Critical points present themselves in many different ways. If $V(x, y, z)$ is a Newtonian potential function the critical points of V are the points where the gravitational force is null. In such examples critical points appear as points of equilibrium. They also appear in problems involving the equilibria of floating bodies, and more generally in any problem in dynamics where there is a generalization of potential.

The following theorem in geography is an old and interesting example of critical point theory. Suppose the surface of the earth is such that the level points are finite in number. Let the points of minimum and maximum distance from the center of the earth be respectively p_0 and p_2 in number. Let the number of passes or saddle points be p_1. If these saddle points are of the "general" non-degenerate type, then

$$p_0 - p_1 + p_2 = 2. \tag{0.1}$$

This example will suggest the following questions. Is (0.1) the only relation between the numbers p_i involved? What happens in n-dimensional spaces? What becomes of the theorem when the number of critical points is infinite? And still more generally when one turns to the critical points of functionals, for example to extremals in boundary problems in the calculus of variations, what becomes of the classification into types of critical points?

As in the study of functions in general, the theory depends in its superficial form upon the class of functions presented. These functions may be analytic, finitely differentiable, continuous, or even semi-continuous. The answer to the first of the preceding questions may be given by presenting a general theorem in the so-called non-degenerate case.[1]

Suppose that f is a single-valued function of a point p on a general n-dimensional manifold R with local coordinates (x). Cf. §10. We suppose that f is of class C^2 in terms of the coordinates (x). We assume that f is *non-degenerate*, that is that the Hessian of f vanishes at no critical point of f. We assume that R is compact. It follows that the number of critical points of f is finite. Let

[1] Morse, M. and van Schaack, G. B., *The critical point theory under general boundary conditions*, Annuals of Mathematics, 35 (1934).

(a) be a critical point of f. By the *index* of (a) is meant the *index*[2] of the quadratic form Q whose coefficients form the Hessian of f at (a). The index is the number of minus signs in the usual canonical form for Q. We see that there are $n + 1$ possible values for the index. This is the simplest form of classification of critical points. It corresponds in dynamics to various degrees of instability at the point of equilibrium. With this understood we state the following theorem.

THEOREM 1. *The numbers p_i of critical points of index i of the non-degenerate function f, and the connectivities R_i of R satisfy the following relations*:

$$
\begin{aligned}
p_0 &\geqq R_0 \\
p_1 - p_0 &\geqq R_1 - R_0 \\
p_2 - p_1 + p_0 &\geqq R_2 - R_1 + R_0 \qquad (0.2)\\
&\cdots\cdots\cdots\cdots \\
p_n - p_{n-1} + \cdots + (-1)^n p_0 &= R_n - R_{n-1} + \cdots + (-1)^n R_0 .
\end{aligned}
$$

Upon considering the successive relations in (0.2) one finds that

$$p_i \geqq R_i \qquad (i = 1, \cdots, n). \tag{0.3}$$

Relations (0.3) imply the existence of what we may call the topologically necessary critical points. If, for example, R were a torus the relations (0.3) would imply that $p_0 \geqq 1$, $p_1 \geqq 2$, and $p_2 \geqq 1$. The relations (0.2) however contain much more. They imply relations between the numbers p_i in excess of those which are topologically necessary. If one is dealing with the problem under general boundary conditions there is a sense in which the preceding relations are complete. See M1, p. 145. We shall leave the non-degenerate problem for the present.

The analogue of the critical point in the calculus of variations is the extremal satisfying given boundary conditions, for example joining two points. The definition and evaluation of the corresponding indices of these critical points have been taken up in M1. Studies of this sort are not the main object of this memoir.

Instead we recall that the theory of functions and functionals can be unified by the study of functions f of a point on an abstract space. Sufficient generality will be obtained if the functions f are lower semi-continuous and the space is a general metric space. See §1. The curves admitted in the calculus of variations become points in this abstract space. The theory as applied to functions and to functionals would still be fairly divergent were it not for the fact that it is possible to introduce a topological definition of a critical point. In the special case of a function which is continuous in the neighborhood of a point (a) of a coordinate space (x), this definition is as follows: *The point (a) is non-critical if there exists a continuous deformation of the domain $f \leqq f(a)$ neighboring*

[2] See Morse, M., *The calculus of variations in the large*, American Mathematical Society Colloquium Publications (1934). We shall refer to these lectures as M1.

(a) under which f is decreased whenever a point p is displaced, and under which the point (a) is actually displaced.

By far the greatest difficulty and the greatest interest arise when one admits the cases where the number of critical values is not finite. The corresponding numbers m_i and R_i will then in general be infinite. They can be defined as cardinal numbers, but the real generalization lies deeper. It depends upon the fact that the entities p_i are merely dimensions of groups M_i and that these groups can be compared at most superficially by comparing their dimensions. The real comparisons are by means of isomorphisms termed "natural" because they are limited in character by the nature of f. In this part of the theory we shall find that the relations (0.2) are replaced by an infinite set of relations of the same formal character, with the numbers p_i and R_i replaced by groups, and the signs $+$ and $-$ replaced by suitably defined operations in abelian group theory of the nature of reductions with respect to a group modulus. The relations (0.2) in the non-degenerate case are a very special case. Even in the non-degenerate case the group theory relations contain new general results.

An extensive bibliography cannot be given here. The reader may however turn to M1 and to the papers[3] cited below for further references. References will be found in M1 to Birkhoff, Brown, Fréchet, Lefschetz, Hadamard, Menger, Poincaré, Schnirelman and Lusternik, Tonelli, and others.

The results here presented form a part of a set of lectures given at the Institute for Advanced Study at Princeton during the year 1935–36. Among those who were present and made helpful suggestions were Drs. Reinhold Baer and E. Čech. See §5. The author's assistant, Dr. Pitcher, rendered valuable service in the reading and preparation of the manuscript.

I. The Underlying Topology and Group Theory

1. The space M and function F. Let M be a space of elements $p, q, r, \cdots$ in which a number pq is assigned to each ordered pair of points such that

I. $pp = 0$,
II. $pq \neq 0$ if $p \neq q$,
III. $pr \leqq pq + rq$.

[3] Some of the more recent references are as follows:

Birkhoff, G. D. and Hestenes, M., *Generalized minimax principle in the calculus of variations*, Duke Mathematical Journal, 1 (1935), 413–432.

Lefschetz, S., *On critical sets*, Duke Mathematical Journal, 1 (1935), 392–412.

Menger, K., *Untersuchungen über allgemeine Metrik*, Mathematische Annalen, 100 (1928), 75–163; 103 (1930), 466–501; *Metrische Geometrie und Variationsrechnung*, Fundamenta Mathematicae, 25 (1935), 441–458.

Morse, M. and van Schaack, G. B., *Abstract critical sets*, Proceedings of the National Academy of Sciences, 21 (1935), 258–263.

Morse, M. and van Schaack, G. B., *Critical point theory under general boundary conditions*, Duke Mathematical Journal, 2 (1936).

Morse, M., *Functional topology and abstract variational theory*, Proceedings of the National Academy of Sciences, 22 (1936).

Lusternik, L. and Schnirelman, L., *Méthodes topologiques dans les problèmes variationnels*, Paris, Hermann & C[ie].

Upon setting $r = p$ in III we see that $pq \geqq 0$ and upon setting $p = q$ that $qr = rq$. The space M is termed a metric space. It is called symmetric since $qr = rq$. The elements p, q, r, etc. are termed "points" and pq the "distance" from p to q.

Neighborhoods, limit points, sets relatively open or closed, can now be defined in the usual way.[4] In particular if e is a positive number the e-neighborhood A_e of a point set A shall consist of all points p with a distance from A less than e. A set $B \subset M$ will be said to be *compact* if every infinite sequence in A contains a Cauchy subsequence which converges to a point in A. We do not assume that the space M is compact.

Let $f(p)$ be a real single-valued function of the point p on a subset A of M. We include $\pm\infty$ among the values which f may assume. We say that f is *lower semi-continuous* at a point q of A if corresponding to each constant $c < f(q)$, there exists an e-neighborhood q_e of q such that $f(p) > c$ on the intersection $A \cdot q_e$ of A and q_e. Upper semi-continuity of f at q is similarly defined, reversing the preceding inequalities. We say that $f(p)$ is continuous at q if its value at q is finite and $f(p)$ is both upper and lower semi-continuous at q.

We shall consider a function $F(p)$ lower semi-continuous on M. We shall suppose that

$$0 \leqq F \leqq 1 \tag{1.1}$$

and that for any constant c such that $0 \leqq c < 1$ the domain $F \leqq c$ is compact.

The functions f which we shall encounter in our abstract variational theory will be bounded below and lower semi-continuous on M. They may take on the value $+\infty$. For each finite constant b the domain $f \leqq b$ will be seen to be compact. Such functions f lead to functions F as follows. Let m be the absolute minimum of f on M. Set $\varphi = f - m$ and

$$F_1 = \frac{\varphi}{1 + \varphi} \tag{1.2}$$

understanding that $F_1 = 1$ when $\varphi = +\infty$. The function F_1 defined by (1.2) is readily seen to have the properties ascribed to the function F.

The set of points at which $F < a$ or $F \leqq a$ will be denoted respectively by F_a or F_{a+}. If A is an arbitrary point set of M we shall say that A is "*definitely on*" F_a (written $\text{on}_\circ\ F_a$) if F is less than a constant $c < a$ at points of A. We note that A may be on F_a without being $\text{on}_\circ\ F_a$ even when A is closed. To illustrate this consider the following example.

Example 1.1. Let M be the curve

$$x^2 + (y - \tfrac{1}{2})^2 = \tfrac{1}{4} \tag{1.3}$$

and suppose that $F = y$ at each point (x, y) of M except at the point $(0, 1)$. At $(0, 1)$ let $F = 0$. The space M is on $F < 1$ but not $\text{on}_\circ\ F < 1$.

---

[4] See Hausdorff, F., *Mengenlehre*, Leipzig, Walter de Gruyter (1927).

2. Vietoris cycles. We shall use Vietoris[5] cycles in our topology. Singular cycles in the classical sense are inadequate in a number of ways. We shall illustrate this fact by an example. Let V be a compact subset of M and V_e the e-neighborhood of V. Let u be an arbitrary singular k-cycle not on V. Corresponding to each positive e we suppose that u is homologous to a cycle on V_e. It is not always true that u is homologous to a cycle on V as examples would show. The corresponding theorem for Vietoris cycles however is true.

We proceed with a systematic outline of the Vietoris theory generalized and modified to meet our needs.

Let (A) be a set of $k + 1$ points of M. For $k > 0$ the orderings of (A) will be divided into two classes, any ordering of one class being obtained from any other of the same class by an even permutation. The set (A) taken with one of these classes of orderings will be termed a positively oriented k-cell and taken with the other class a negatively oriented k-cell. An oriented k-cell α_k may be represented by a succession

$$\eta\, A_0 \cdots A_k \tag{2.0}$$

of its vertices preceded by η, where $\eta = 1$ or -1 according as the ordering $A_0 \cdots A_k$ belongs to α_k or not. In case $k = 0$ we use (2.0) to define an oriented 0-cell understanding the cell is positively or negatively oriented according as η is 1 or -1. We say that the oriented $(k - 1)$-cell

$$\eta(-1)^i A_0 \cdots A_{i-1} \cdot A_{i+1} \cdots A_k$$

is positively related to α_k. We shall say that α_k is of *norm* e if the distances between the vertices of α_k are less than e.

The ce'l α_k will be termed *degenerate* or *null* if at least two of its vertices are coincident. Let δ be an element in an arbitrary field[6] Δ. By a k-chain of norm e is meant a symbolic[7] sum Σ of the form $\delta^i a^i$, $i = 1, \cdots, m$, in which δ^i is an element of Δ and a^i is an oriented k-cell of norm e. The chain Σ will be termed *reduced* if none of the a^i are null, if each is positively oriented, and no two of the a^i are identical.

An arbitrary chain will be *reduced* as follows. Let a be an arbitrary positively oriented k-cell and b the corresponding negatively oriented k-cell. Any term of the form δb in Σ will be replaced by $-\delta a$. All terms involving a will be replaced by a single term $\delta' a$ in which δ' is the sum of the coefficients δ in the terms replaced. Finally all terms involving null cells or null coefficients will be dropped. The resulting reduced chain will be regarded as formally equal to the original chain Σ.

---

[5] Vietoris, L., *Über den höheren Zusammenhang kompakter Räume und eine Klasse von zusammenhangstreuen Abbildungen*, Mathematische Annalen, 97 (1927), 454–472.

See also Lefschetz, S., *Topology*, American Mathematical Society Colloquium Publications, New York (1930); and *Deformations*, Duke Mathematical Journal, 1 (1935).

[6] See van der Waerden, B. L., *Moderne Algebra*, p. 41, Berlin, Julius Springer (1930).

[7] We adopt the summation conventions of tensor analysis.

By the sum of two chains $\delta^i a^i$ and $\delta^{i'} a^{i'}$ is meant the chain $\delta^i a^i + \delta^{i'} a^{i'}$. We assume that

$$\delta[\delta^i x^i] = (\delta\delta^i)x^i.$$

It is readily seen that k-chains of norm e form an additive abelian operator group.[8]

We shall now define the boundary operator β. If a is an arbitrary oriented k-cell with $k > 0$, and e is the unit element in Δ, ea will be a k-chain and $\beta(ea)$ shall be the $(k-1)$-chain $\sum_i eb_i$ where b_i is an arbitrary $(k-1)$-cell positively related to a. More generally we set

$$\beta(\delta^i a^i) = \delta^i[\beta(ea^i)].$$

If $k = 0$ we understand that $\beta(ea) = 0$. If u is an arbitrary k-chain one sees[9] that $\beta\beta u = 0$. One terms βu the *boundary* chain defined by u.

The preceding k-chains are termed *algebraic* k-chains to distinguish them from Vietoris chains to be defined later. The term algebraic will be omitted when it is clear from the context that the chain is algebraic. In particular this will be the case whenever the norm e of the chain is mentioned.

Let B and C be compact subsets of M such that $B \subset C$. An algebraic k-chain u will be termed a cycle *modulo* B on C if βu is on B. We term u an absolute k-cycle or a relative k-cycle according as B is or is not null. The adjective absolute will ordinarily be omitted. We shall term u, e-homologous to 0 *modulo* B on C and write

$$u \sim_e 0 \qquad (\text{mod } B \text{ on } C)$$

if there exists a $(k+1)$-chain z of norm e on C such that $\beta z = u + v$, where v is a k-chain of norm e on B. If $B = 0$, then $v = 0$ and the phrase mod B is omitted.

Let $u = (u_n)$ be a sequence of algebraic k-cycles u_n, $n = 0, \cdots$, mod B on C, with norms at most e_n respectively. If the numbers e_n tend to zero as n becomes infinite, and if for each integer n there exist "connecting" homologies of the form

$$u_n \sim_{e_n} u_{n+1} \qquad (\text{mod } B \text{ on } C), \tag{2.1}$$

u is termed a (Vietoris) k-cycle mod B on C, and C a *carrier* of u. We term u homologous to zero mod B on C, and write $u \sim 0$ mod B on C if corresponding to each positive number e there exists an integer N so large that $u_n \sim_e 0$ mod B on C whenever $n > N$. The set C is termed a *carrier* of the homology $u \sim 0$ mod B on C. Vietoris k-cycles u and v are termed homologous, $u \sim v$, mod B on C, if $u - v \sim 0$ mod B on C. (See below for definition of $u - v$.)

The algebraic k-cycles u_n will be termed the *components* of $u = (u_n)$. If u and v are Vietoris k-cycles mod B on C the algebraic k-cycles $u_n \pm v_n$ are the components of Vietoris k-cycle which we denote by $u \pm v$. If δ is an ele-

[8] See van der Waerden, l.c. pp. 132–144.

[9] See Seifert-Threlfall, *Lehrbuch der Topologie*, p. 60, Leipzig, Teubner (1934).

ment in the field Δ the algebraic k-cycles δu_n are the components of Vietoris k-cycle which we denote by δu.

In the remainder of this memoir the term k-cycle mod B on C shall mean a Vietoris k-cycle mod B on C unless otherwise stated.

If (u_n) is a k-cycle mod B on C any infinite subsequence (v_n) of (u_n) defines a k-cycle $v \sim u$ mod B on C. With this understood let ζ be an algebraic k-cycle mod B on C of norm e such that $u_n \sim_e \zeta$ mod B on C for all integers n exceeding some integer N. We then write

$$u \sim_e \zeta \qquad \text{(mod } B \text{ on } C). \tag{2.2}$$

By virtue of these conventions we have the following lemma.

LEMMA 2.1. *If $u = (u_n)$ and $v = (v_n)$ are k-cycles mod B on C such that for every n,*

$$u_n \sim_{e_n} v_n \qquad \text{(mod } B \text{ on } C) \tag{2.3}$$

where e_n tends to zero as n becomes infinite, then $u \sim v$ mod B on C.

In the preceding paragraphs we have spoken of algebraic and Vietoris k-cycles and homologies mod B on C where B and C were compact. If the phrase mod B on C is replaced by mod B' on C' where B' and C' are compact or not at pleasure, mod B' on C' shall mean mod B'' on C'', where B'' and C'' are respectively compact sets on B' and C'.

Letting X_e denote the e-neighborhood of X we state the following lemma.

LEMMA 2.2. *If $u = (u_n)$ is a k-cycle mod B on C such that $u \sim 0$ mod B_e on C_e for every positive e, then $u \sim 0$ mod B on C.*

Corresponding to an arbitrary positive e there exists by hypothesis an integer N so large that for $n > N$, $u_n \sim_e 0$ mod B_e on C_e. For such values of n we infer the existence of an algebraic $(k+1)$-chain w_n on C_e of norm e and an algebraic k-chain v_n on B_e of norm e such that

$$\beta w_n = u_n + v_n .$$

Let v_n' be the algebraic k-chain obtained by replacing each vertex of v_n not on B by a nearest point on B. It is easily seen that there exists an algebraic $(k+1)$-chain z_n on B_e of norm $3e$ such that $\beta z_n = v_n' - v_n$. If we set $w_n' = w_n + z_n$ we find that

$$\beta w_n' = u_n + v_n' .$$

We now replace each vertex of w_n' on C_e but not on C by a nearest vertex on C obtaining thereby a $(k+1)$-chain y_n on C of norm at most $5e$. We have the relation

$$\beta y_n = u_n + v_n'$$

and since v_n' is on B, $u \sim 0$ mod B on C.

3. **Theorems on Vietoris k-cycles.** Let g be an arbitrary additive abelian operator group with coefficients in a field Δ. With such a group there can be associated in many ways a *maximal linear set* of elements of g, that is a set S

of elements of g no proper linear combination of which with coefficients in Δ is null, and which is a proper subset of no other such set. The cardinal number m of the set S equals that of any other maximal linear set of g. This is readily proved in case m is finite. It is true even when m is infinite but we shall make no use of this fact. We term m the *dimension* of g.

Let V be a compact subset of M. We are dealing with Vietoris k-cycles. If u were an ordinary singular k-cycle homologous to zero mod V it is clear that u would be homologous to a singular k-cycle on V. If however u is a Vietoris k-cycle the corresponding theorem requires proof. Theorems 3.1, 3.2, and 3.3 are general theorems of this nature. If the field Δ consisted merely of the integers mod 2, use could be made of the compactness of the relative homology groups to prove these theorems. But with our general group Δ the proof is more difficult.

The basic ideas appear in two lemmas whose proof was communicated to the author in slightly different form by Professor E. Čech. Professor Čech had previously developed similar lemmas and theorems in his abstract theory of spaces.[10]

We begin with a definition.

Reduction sets W. Let $U \subset V$ be compact subsets of M. Corresponding to each positive δ let $W(\delta)$ be a group of algebraic k-cycles mod U on V of norm δ, with the property that for each positive $\eta < \delta$, $W(\eta)$ is a subgroup of $W(\delta)$. Such a set of groups $W(\delta)$ will be called a *reduction set W*.

The first lemma here is as follows.

LEMMA 3.1. *Corresponding to an arbitrary positive constant e there exists a positive constant $\delta < e$ such that to each cycle $w(\delta)$ in $W(\delta)$ there corresponds for every $\eta < \delta$ at least one cycle $w(\eta)$ in $W(\eta)$ such that*

$$w(\delta) \sim_e w(\eta) \qquad (\text{mod } U \text{ on } V). \tag{3.1}$$

Let $e_n = e\,3^{-n}$. Let ω denote the subgroup of cycles of $W(e_1)$ which are e-homologous to zero mod U on V. Let h_1 denote the group $W(e_1)$ mod ω. We shall begin with a proof of statement (α).

(α). The dimension of h_1 is finite.[11]

Let A be a finite net of points on U within a distance e_1 of each point of U. Let $B \supset A$ be a similar net for V. Let g denote the group of algebraic k-cycles mod A on B. It is clear that the dimension of g is finite.

Corresponding to an arbitrary algebraic k-cycle $w = w(e_1)$ of $W(e_1)$ we can obtain an algebraic k-cycle u of g of norm $e = 3e_1$ by replacing each vertex of w on U by a nearest point of A and each vertex of w not on U by a nearest point of B. We infer that $w \sim_e u$ mod U on V. But the group g admits a finite base with elements z^i so that $u = \delta^i z^i$ and hence

$$w \sim_e \delta^i z^i \qquad (\text{mod } U \text{ on } V).$$

[10] Čech, E., *Théorie générale de l'homologie dans un espace quelconque*, Fundamenta Mathematicae, 19 (1932), 149–183.

[11] Cf. Vietoris, l.c., p. 461 (3).

It follows that the dimension of h_1 is at most the number of elements z^i, and hence is finite.

The proof of (α) is complete.

Recall that h_1 is a group of classes of k-cycles. For each integer $n > 1$ let h_n be the subgroup of those classes of h_1 which contain at least one cycle of $W(e_n)$. We see that

$$h_1 \supset h_2 \supset h_3 \supset \cdots .$$

There must accordingly exist a finite integer r such that

$$\dim h_r = \dim h_{r+1} = \cdots . \tag{3.2}$$

But two operator groups with coefficients in a field and with equal finite dimensions will be identical if one group is a subgroup of the other. Hence

$$h_r = h_{r+1} = \cdots . \tag{3.3}$$

The constant e was arbitrary and $e_n = e\,3^{-n}$. We now let δ be any positive constant such that $\delta < e_r$. The cycle $w(\delta)$ of the lemma is in $W(e_r)$ and hence in some class of h_r. Corresponding to the constant $\eta < \delta$ of the lemma let p be an integer so large that $e_p < \eta$. Then $e_p < \delta < e_r$ so that $p > r$ and $h_p = h_r$. The cycle $w(\delta)$ is in a class of h_r and hence of h_p, and by virtue of its definition this class of h_p contains at least one k-cycle $w(e_p)$ of $W(e_p)$. Since $e_p < \eta$ the cycle $w(e_p)$ is a cycle $w(\eta)$. The cycles $w(\delta)$ and $w(\eta)$ are in the same class of h_p and hence of h_1. The relation (3.1) is satisfied by virtue of the definition of h_1.

The proof of the lemma is complete.

We now define ensembles Z closely related to the reduction sets W.

Ensembles Z. Corresponding to each positive δ let $Z(\delta)$ be a non-null class of algebraic k-cycles $z(\delta)$ mod U on V of norm δ. The ensemble of the classes $Z(\delta)$ as δ ranges over all positive values will be denoted by Z and assumed to have the following properties:

(1). For an arbitrary $\eta < \delta$, $Z(\eta)$ shall be a subset of $Z(\delta)$.

(2). With Z there shall be associated a reduction set W such that if $z(\delta)$ is an arbitrary element of $Z(\delta)$ any other element of $Z(\delta)$ is of the form

$$z(\delta) + w(\delta) \qquad [w(\delta) \subset W(\delta)], \tag{3.4}$$

and every sum of the form (3.4) is an element of $Z(\delta)$.

We shall illustrate the ensembles Z and reduction sets W by the following example.

Example 3.1. Let $U \subset V \subset C$ be compact subsets of M. Let u be a Vietoris k-cycle mod U on C homologous to zero mod V on C. Let $Z(\delta)$ and $W(\delta)$ consist respectively of the algebraic k-cycles $z(\delta)$ and $w(\delta)$ mod U on V of norm δ such that

$$z(\delta) \sim_\delta u, \qquad w(\delta) \sim_\delta 0 \qquad (\text{mod } U \text{ on } C). \tag{3.5}$$

That there exist algebraic k-cycles $z(\delta)$ in (3.5) may be seen as follows. Since $u \sim 0$ mod V on C there exists a sequence of positive constants e_n tending to zero, and algebraic chains u_{n+1} of norm e_n on C with algebraic k-chains z_n of norm e_n on V such that

$$\beta u_{n+1} = u_n - z_n.$$

Hence $\beta u_n = \beta z_n$. It follows that z_n is an algebraic k-cycle mod U on V and that

$$z_n \sim_{e_n} u_n \qquad (\text{mod } U \text{ on } C).$$

The classes $Z(\delta)$ are accordingly not empty.

The remaining conditions on ensembles Z and reduction sets W are easily seen to be satisfied.

We shall now prove the following lemma.

LEMMA 3.2. *Corresponding to an admissible ensemble Z there exists a Vietoris k-cycle (v_n) mod U on V whose components v_n belong to $Z(e_n)$ for suitable positive constants e_n approaching zero as n becomes infinite.*

We recall that there is a reduction set W associated with Z. Beginning with an arbitrary positive number e_0 we shall give an inductive definition of a sequence of positive numbers e_n tending to zero as n becomes infinite. The constant e of Lemma 3.1 is arbitrary. In particular we can set $e = e_{n-1}$. Lemma 3.1 then affirms the existence of a constant $\delta < e$. We take $e_n < \delta$. We suppose also that e_n tends to zero as n becomes infinite.

By definition of Z there is an algebraic k-cycle z_n mod U on V in $Z(e_n)$ for each n. We shall give an inductive definition of a Vietoris k-cycle $v = (v_n)$ mod U on V. We begin by setting $v_1 = z_1$. Suppose that the components v_i, $i = 1, \cdots, m$, have been defined in such fashion that v_i is in $Z(e_i)$ and

$$v_{i+1} \sim_{e_{i-1}} v_i \qquad (\text{mod } U \text{ on } V;\ i = 1, \cdots, m-1).$$

We shall now define v_{m+1}. First recall that $z_{m+1} - v_m$ is in $W(e_m)$ in accordance with property (2) of ensembles Z. We can apply Lemma 3.1 with $e = e_{m-1}$, with $e_m < \delta < e$ and with

$$w(\delta) = z_{m+1} - v_m.$$

Setting $\eta = e_{m+1}$ Lemma 3.1 affirms the existence of a cycle $w(\eta)$ in $W(e_{m+1})$ such that

$$w(\delta) \sim_{e_{m-1}} w(\eta) \qquad (\text{mod } U \text{ on } V). \tag{3.6}$$

We set

$$v_{m+1} = z_{m+1} - w(\eta). \tag{3.7}$$

We note that v_{m+1} is in $Z(e_{m+1})$ by virtue of property (2) of ensembles Z. Upon adding the respective members of (3.6) and (3.7) we find that

$$v_{m+1} \sim_{e_{m-1}} v_m \qquad (\text{mod } U \text{ on } V).$$

The algebraic k-cycles v_n are accordingly the components of a Vietoris k-cycle v.

The proof of the lemma is complete.

We shall prove three theorems which parallel facts obvious in the theory of singular cycles. Letting $U \subset V \subset C$ be three compact subsets of M we state the first theorem as follows.

THEOREM 3.1. *If u is a k-cycle mod U on C homologous to zero mod V on C then u is homologous mod U on C to a k-cycle z mod U on V.*

This theorem follows at once from Lemma 3.2 with the choice of the ensemble Z that of Example 3.1.

If w and u are two Vietoris k-cycles we shall find it convenient at times to write $w \equiv u \bmod A$. We mean thereby that $w_n = u_n + v_n$ where v_n is a chain on A and $n = 1, 2, \cdots$.

Letting $A \subset U \subset V$ be three compact subsets of M the second theorem is as follows.

THEOREM 3.2. *If a $(k-1)$-cycle u mod A on U is homologous to zero mod A on V there exists a k-cycle v mod U on V such that $\beta v \equiv u$ mod A.*

This theorem also follows from Lemma 3.2 We take $Z(\delta)$ as the class of all k-cycles $z(\delta)$ mod U on V of norm δ such that

$$\beta z(\delta) \sim_\delta u \qquad (\text{mod } A \text{ on } U).$$

The cycles $z(\delta)$ exist by virtue of the hypothesis of the theorem. The corresponding reduction group $W(\delta)$ consists of all k-cycles $w(\delta)$ mod U on V of norm δ such that

$$\beta w(\delta) \sim_\delta 0 \qquad (\text{mod } A \text{ on } U).$$

With this choice of Z it follows from Lemma 3.2 that there exists a k-cycle z mod U on V such that $\beta z \sim u$ mod A on U. There accordingly exists a sequence of positive constants e_n tending to zero, and algebraic $(k+1)$-chains w_n of norm e_n on U with algebraic k-chains x_n of norm e_n on A such that

$$\beta w_n = u_n - \beta z_n + x_n$$

If we set $v_n = w_n + z_n$ we see that (v_n) is a Vietoris k-cycle v mod U on V and that

$$\beta v \equiv u \qquad (\text{mod } A).$$

The proof of the theorem is complete.

The third theorem is as follows.

THEOREM 3.3. *If u is a k-cycle mod A on V and $\beta u \sim 0$ on A, there exists a k-cycle v on V such that $u \sim v$ mod A on V.*

Before proving this theorem it will be convenient to establish the following.

(α). *If x is an algebraic k-cycle x of norm e mod A on V, and $\beta x \sim_e 0$ on A, there exists an algebraic k-cycle y on V such that $x \sim_e y$ mod A on V.*

It follows from the hypothesis in (α) that there exists an algebraic k-chain ω

of norm e on A such that $\beta\omega = \beta x$. We set $y = x - \omega$ and note that y is a k-cycle on V of norm e. To establish (α) we observe that

$$\beta 0 = -\omega + \omega = (y - x) + \omega,$$

so that $x \sim_e y$ mod A on V.

To prove the theorem we shall apply Lemma 3.2 setting $U = 0$. The set $Z(\delta)$ shall be the class of all k-cycles $z(\delta)$ of norm δ on V such that $z(\delta) \sim_\delta u$ mod A on V. The corresponding group $W(\delta)$ will consist of all k-cycles $w(\delta)$ of norm δ on V such that $w(\delta) \sim_\delta 0$ mod A on V. That the set $Z(\delta)$ is not empty follows from (α), and the theorem follows from Lemma 3.2.

4. **The rank conditions.** In this section we shall define an F-non-bounding k-cycle and assign to such a cycle a rank r. These ranks are fundamental. They lead to a group theoretic formulation of the structure beneath our theory. They are introduced here for the first time.

Recall that the sets of points at which $F < c$ and $F \leqq c$ are respectively denoted by F_c and F_{c+}. The phrase on_o F_c of §1 will sometimes be replaced by the phrase below_o c. Mod_o F_c shall mean mod some compact set below_o c.

If z is a k-cycle below_o b, not homologous to zero below_o b we term b a *homology limit* of z, and z F-non-bounding. By the *superior cycle limit* $s(z)$ of an F-non-bounding k-cycle z is meant the least upper bound of the homology limits of z. If z is non-bounding on M we set $s(z) = +\infty$. It follows from the definitions involved that $s(z)$ is a homology limit of z.

Let z be a k-cycle with superior cycle limit $s(z) = s$. By the *inferior cycle limit* $t = t(z)$ of z is meant the greatest lower bound of constants c such that $z \sim 0$ mod_o F_c below_o s. We do not define $t(0)$. We observe that $z \nsim 0$ mod_o F_t below_o s.

If z is an F-non-bounding k-cycle we term the pair $[s(z), t(z)]$ the *rank* $r(z)$ of z. We order the pairs $[s, t]$ lexicographically. That is, we write $[s, t] = [s', t']$ if and only if $s = s'$ and $t = t'$, and write $(s, t) > (s', t')$ if $s > s'$, or if $s = s'$ and $t > t'$. In this way we are led to a simply ordered set of ranks r. These ranks are fundamental.

A bounding k-cycle which is F-non-bounding will be termed *ambiguous*. The superior cycle limit of such a cycle will be finite.

In §5 we shall give the theory of ranks an abstract formulation and we have there listed four conditions which we term *rank conditions*. These rank conditions are satisfied by the ranks $r(u)$, by the superior cycle limits $s(u)$, and by several other basic ordered sets. Our first theorem in this connection is the following.

THEOREM 4.1. *The superior cycle limits $s(u)$ of F-non-bounding k-cycles satisfy the rank conditions* I, II, III, IV *with $\rho(u)$ replaced by $s(u)$.*

I. To show that $s(\delta u) = s(u)$ for $\delta \neq 0$ we must show that the condition $u \sim 0$ below_o b is equivalent to the condition $\delta u \sim 0$ below_o b. The first condi-

tion clearly implies the second. But the second condition also implies the first. For any coefficient in the field Δ can be divided by a non-null δ. In particular if $\delta \neq 0$ the coefficients in any sequence of algebraic bounding relations which imply that $\delta u \sim 0$, can be divided by δ. Thus I holds as stated.

II. To establish II we suppose that $s(u)$ and $s(v)$ are both finite. The validity of II when $s(u)$ or $s(v)$ is infinite is trivial. Let e be an arbitrary positive constant. By virtue of the definition of a superior cycle limit, $u \sim 0$ below$_\circ$ $s(u) + e$ and $v \sim 0$ below$_\circ$ $s(v) + e$. Hence $u + v \sim 0$ below$_\circ$ $s^* + e$ where s^* is the maximum of $s(u)$ and $s(v)$. Property II follows as stated.

III. To establish III we suppose that $s(u) > s(v)$ and let b be any homology limit of u between $s(u)$ and $s(v)$. Observe that $v \sim 0$ and $u \nsim 0$ below$_\circ$ b so that $u + v \nsim 0$ below$_\circ$ b. Thus b is a homology limit of $u + v$, and III holds as stated.

IV. We set $\rho_0 = s_0$. Observe that u and v in IV are below$_\circ$ s_0. Since $s(u)$ and $s(v)$ do not exist $u \sim 0$ and $v \sim 0$ below$_\circ$ s_0. If $s(u + v)$ exists it must be less than s_0, and IV is established.

The proof of Theorem 4.1 is complete.

It might be thought that the inferior cycle limits $t(u)$ also satisfy the rank axioms. To show that this is not the case and to illustrate the preceding theorems we present several examples.

Let the metric space M be the curve

$$y = -e^{-x} \sin x \qquad (0 \leqq x \leqq 4\pi) \tag{4.1}$$

in the (x, y) plane. On M let $F = y$. The function F is continuous on M. We waive the condition $0 \leqq F \leqq 1$. We shall denote the points of M at which $x = 0, \pi, 2\pi$, and 4π by a, b, c, e respectively, and the points at which x equals

$$\frac{5\pi}{4}, \quad \frac{9\pi}{4}, \quad \frac{13\pi}{4}$$

respectively by n, p, q. The reader will do well to trace the curve M and note that n, p, q are relative extrema of F. We observe that

$$F(p) < F(q) < F(n).$$

Let A, B, C, E be 0-cycles mod 2 whose components are respectively identical with a, b, c, e.

Example 4.1. The cycle $B - C$ has the superior cycle limit $F(n)$ while $B - A$ has no homology limit.

Example 4.2. In this example we shall show that cycles u and v may have homology limits while $u + v$ has no homology limit. To that end let

$$u = C - A, \qquad v = B - C, \qquad u + v = B - A.$$

We observe that $s(u) = s(v) = F(n)$ while $s(u + v)$ does not exist.

Example 4.3. We here show that the equality is possible in rank axiom II as applied to s. We set

$$u = B - C, \qquad v = C - E,$$

and note that $u + v = B - E$. We find that

$$s(u) = s(u + v) = F(n), \qquad s(v) = F(q) < s(u).$$

Example 4.4. We shall show that the sign $<$ is possible in rank axiom II. To this end we make use of the cycles u and v in Example 4.3 and set $u' = u + v$, $v' = -u$. We observe that $u' + v' = v$. Referring to Example 4.3 we find that

$$s(u' + v') < \max [s(u'), s(v')].$$

In the next three examples we shall show that for inferior cycle limits each one of the three cases

$$t(u + v) \gtreqless \max [t(u), t(v)] \tag{4.2}$$

is possible.

Example 4.5. We shall first show the possibility of the sign $>$ in (4.2). We make use of the space M and function F of the preceding examples and set $u = B - C$, $v = E - B$. We observe that $u + v = E - C$ and that

$$s(u) = s(v) = F(n), \qquad s(u + v) = F(q),$$

$$t(u) = t(v) = F(p), \qquad t(u + v) = 0 > t(u) = t(v).$$

Example 4.6. We here illustrate the possibility of the equality in (4.2). Let the space M consist of two points a and b. Let the function F be 1 at a and 0 at b. Let u and v be two 0-cycles mod 2 with components respectively identical with a and b. We find that

$$s(u) = s(v) = s(u + v) = +\infty,$$

$$t(u) = t(u + v) = 1, \qquad t(v) = 0.$$

Example 4.7. We shall show finally that the sign $<$ is possible in (4.2). We make use of the space M and function F of the preceding example, and set $z = v - u$. We find that

$$t(z) = t(u) = 1, \qquad t(u + z) = 0.$$

We return to the general theory and state a basic theorem.

THEOREM 4.2. *The ranks $r(u)$ of F-non-bounding k-cycles satisfy the four rank conditions of* §5, *$r(u)$ replacing $\rho(u)$.*

Let b be a homology limit of a k-cycle u. Let $I[u, b]$ denote the greatest lower bound of constants c such that $u \sim 0 \bmod_o F_c$ on$_o$ F_b. We state the following lemma.

LEMMA 4.1. *If b is a homology limit of u, v and $u + v$, then*

$$I[u + v, b] \leqq \max [I(u, b), I(v, b)]. \tag{4.3}$$

If b is a homology limit of u and of v and if

$$I(u, b) > I(v, b),$$

then b is a homology limit of $u + v$ and

$$I(u + v, b) = I(u, b).$$

The lemma follows from the definitions involved.

We shall establish Theorem 4.2.

I. That $t(\delta u) = t(u)$ if $\delta \neq 0$ follows as in the proof that $s(\delta u) = s(u)$. Hence $r(\delta u) = r(u)$.

II. To establish II we may assume without loss of generality that $r(v) \leqq r(u)$. We have then to prove that $r(u + v) \leqq r(u)$. According to Theorem 4.1 we have two cases:

(a) $s(u + v) < s(u)$.

(b) $s(u + v) = s(u)$.

In case (a), $r(u + v) < r(u)$ by virtue of the ordering of pairs (s, t). In case (b) we distinguish between the two subcases:

(b)′ $s(v) < s(u)$.

(b)″ $s(u) = s(v)$.

In case (b)′ we observe that $v \sim 0$ below$_\circ$ $s(u)$ so that $u + v \sim u$ below$_\circ$ $s(u)$. Hence $t(u + v) = t(u)$. Thus $r(u + v) = r(u)$ in case (b)′, and II is true.

In case (b)″, u, v, and $u + v$ have common superior cycle limits. It follows from the preceding lemma that

$$t(u + v) \leqq \max [t(u), t(v)]. \tag{4.4}$$

Hence $r(u + v)$ is at most the maximum of $r(u)$ and $r(v)$, and II is established.

III. We here show that $r(u + v)$ exists, provided $r(u)$ and $r(v)$ exist and are unequal. We assume without loss of generality that $r(u) > r(v)$ and consider the two cases:

(α) $s(u) > s(v)$.

(β) $s(u) = s(v)$, $\qquad t(u) > t(v)$.

In case (α) it follows from Theorem 4.1 that $s(u + v)$ exists, and in case (β), $s(u + v)$ exists by virtue of the preceding lemma. Hence $r(u + v)$ exists in both cases, and III is proved.

IV. If the cycles $u = \Sigma u_i$ and $v = \Sigma v_j$ in IV have no rank, $s(u)$ and $s(v)$ fail to exist. But $r(u + v)$ and hence $s(u + v)$ exist by hypothesis. We set $\rho_0 = r_0 = (s_0, t_0)$. It follows from the fact that superior cycle limits satisfy IV that $s(u + v) < s_0$, and hence $r(u + v) < r_0$. Hence IV holds as stated.

The proof of Theorem 4.2 is complete.

5. **The underlying group structure.** The implications of the rank conditions will be best understood if incorporated in an abstract group theory.

We suppose then that we have an additive abelian operator group G with operators which form a field Δ. We shall be concerned with various subgroups g of G. We suppose throughout that these subgroups are operator subgroups, that is if u is in g and δ is in Δ, δu is in g. Our isomorphisms shall be operator isomorphisms, that is, if u corresponds to v, δu corresponds to δv.

By a subgroup g of G with property A is meant an operator subgroup of G every element of which, with the possible exception of the null element, has the property A. A group g with property A will be termed *maximal* if it is a proper subgroup of no other subgroup g' of G with property A. The following example shows that there may be several maximal groups with a given property A.

Let G be a group generated by three elements a, b, c with coefficients in the field of integers mod 2. Suppose that a, b, $a + b$, c are the only elements with the property A. Then a and b together generate a maximal group with property A as does c.

Suppose that the respective elements u of G have the property A if and only if for each operator $\delta \neq 0$, δu has the property A. Let g be a maximal (operator) subgroup of G with property A. Let v be an arbitrary element with property A. There will then exist an operator $\delta \neq 0$ and an element v_1 in g, together with an element v_2 which fails to have the property A, or is null, such that $\delta v = v_1 + v_2$. Now division by δ is permissible, v_1/δ is in g, and v_2/δ still fails to have the property A or is null. We can accordingly infer the existence of an element z in g, and an element w which fails to have the property A or is null, such that

$$v = z + w. \tag{5.1}$$

The rank conditions. Let (ρ) be a simply ordered set of elements, symbols, numbers, points, etc. With some but not all of the elements u of G we suppose there is associated an element $\rho(u)$ in the set (ρ). We term $\rho(u)$ the rank of u. The rank $\rho(0)$ shall not be defined.

We assume that the ranks ρ satisfy the following conditions:

I. *If u has a rank and $\delta \neq 0$, $\rho(u) = \rho(\delta u)$.*

II. *If u, v, and $u + v$ have ranks, $\rho(u + v) \leqq max\ [\rho(u), \rho(v)]$.*

III. *If u and v have unequal ranks, $\rho(u + v)$ exists.*

IV. *If $u_1, \cdots, u_m, v_1, \cdots, v_n$ have ranks at most ρ_0 while the sums $u = \Sigma u_i$ and $v = \Sigma v_j$ have no rank and $u + v$ has a rank, then $\rho(u + v) < \rho_0$.*

The elements with rank (with the null element added) will not in general form a group. For example consider the group G of 0-cycles in Example 4.2 letting the superior cycle limit $s(z)$ be the rank of the 0-cycle z in case $s(z)$ exists. In this example we have seen that $s(u)$ and $s(v)$ exist while $s(u + v)$ does not exist nor does $u + v = 0$.

We have already remarked that the primitive concept of counting is to be

replaced by comparisons involving isomorphisms of groups. A comparison by means of isomorphisms in general would not be very effective. Our comparisons will always be by means of isomorphisms of a special sort the abstract counterpart of which we now define.

Two elements of G will be said to be in the same *ρ-class* if they have the same rank while their difference has no rank or a lesser rank. An isomorphism between two subgroups of G of elements with rank will be termed a *ρ-isomorphism* if corresponding non-null elements are in the same ρ-class. These ρ-isomorphisms will be exemplified by three classes of *natural isomorphisms* defined by F.

We state the following theorem.

THEOREM 5.1. *Any two maximal groups of elements of G with a fixed rank σ are ρ-isomorphic.*

Let g_σ be the group generated by the elements of G with rank at most σ. Let H_σ be the subset of elements of g_σ without rank or with rank less than σ. We continue by proving the following statement.

(α). *The elements of H_σ form a group.*

Let u and v be arbitrary elements of H_σ. Writing E for exists and $\sim E$ for does not exist we distinguish four cases:

(1) $\rho(u)\ E,\ \rho(v)\ E.$ (2) $\rho(u)\ E,\ \rho(v) \sim E.$

(3) $\rho(u) \sim E,\ \rho(v)\ E.$ (4) $\rho(u) \sim E,\ \rho(v) \sim E.$

We shall prove that $u + v$ is in H_σ. If $\rho(u + v)$ does not exist $u + v$ is in H_σ. We assume therefore that $\rho(u + v)$ exists and seek to prove that $\rho(u + v) < \sigma$. This will follow from the rank conditions II, III, III, IV respectively in the cases (1), (2), (3), (4). In case (1), $\rho(u)$ and $\rho(v)$ are less than σ since u and v are in H_σ. It follows from II that $\rho(u + v) < \sigma$. In case (2), $\rho(u) < \sigma$ and $\rho(u + v)$ must be less than σ for otherwise $\rho(v)$ would exist. Case (3) is similar. The result in case (4) follows from IV, and the proof of (α) is complete.

We return to the theorem and let m_σ be a maximal group of elements of G with rank σ. Each element of m_σ is in a coset of the residue group h_σ defined by g_σ mod H_σ, and different elements u and v of m_σ are in different cosets of h_σ, for otherwise $u - v$ would be in H_σ and hence not in m_σ. Moreover there s an element of m_σ in each coset z of h_σ for otherwise an element u of z taken with m_σ would generate a group of elements of G with rank σ containing m_σ as a proper subgroup. There is thus an isomorphism between m_σ and h_σ in which each element u of m_σ corresponds to the coset of h_σ which contains u. We term m_σ and h_σ *ρ-isomorphic* since u corresponds to a coset of elements in the same ρ-class as u.

Let m'_σ be a second maximal group of elements of G with rank σ. The group m'_σ is also ρ-isomorphic with h_σ. It follows that m_σ and m'_σ are ρ-isomorphic with each other, and the proof of the theorem is complete.

A particular consequence of the theorem is that the dimensions of any two maximal groups of elements of G with rank σ are equal, but the theorem of course implies much more than this equality of dimension.

We note another consequence of statement (α).

(β). *The property of elements with rank being in the same ρ-class is transitive.* For if u, v, and w have the rank σ and if u is in the ρ-class of v and w, $u - v$ and $u - w$ belong to H_σ. It follows from (α) that $v - w$ belongs to H_σ so that v and w are in the same ρ-class.

We state the following lemma.

LEMMA 5.1. *If $u_1, \cdots, u_m$ are elements of G with ranks such that*

$$\rho(u_1) > \rho(u_i) \qquad (i = 2, \cdots, m) \tag{5.2}$$

then $\rho(u_1 + \cdots + u_m)$ exists and equals $\rho(u_1)$.

Suppose first that $m = 2$. By virtue of rank condition III $\rho(u_1 + u_2)$ exists. From I we see that $\rho(-u_2) = \rho(u_2)$. Since $u_1 = (u_1 + u_2 - u_2)$ we can infer from II and (5.2) that

$$\rho(u_1) \leqq \max [\rho(u_1 + u_2), \rho(u_2)] \leqq \rho(u_1 + u_2),$$

and that

$$\rho(u_1 + u_2) \leqq \max [\rho(u_1), \rho(u_2)] \leqq \rho(u_1)$$

so that $\rho(u_1) = \rho(u_1 + u_2)$. The proof of the lemma can be completed by induction with respect to m.

Let there be given a subgroup g of G and a set of subgroups $h(\alpha)$ of g, α being an enumerating index. The group g is said to be a *direct sum*

$$g = \sum_\alpha h(\alpha)$$

of the groups $h(\alpha)$ if each element u of g is a finite sum of elements from the groups $h(\alpha)$, and if there exists no relation of the form

$$u_{\alpha_1} + \cdots + u_{\alpha_m} = 0$$

in which the α_i's are distinct and u_{α_i} is a non-null element from the group $h(\alpha_i)$.

From this point on we shall frequently employ the hypothesis that the ranks ρ are well-ordered. We shall not mean thereby that the ranks *can* be well-ordered but that taken in their given orders they *are* well-ordered.[12]

We state the following theorem.

THEOREM 5.2. *If the ranks of elements of G are well-ordered a maximal subgroup g of the elements of G with rank is a direct sum*

$$\sum_\rho h(\rho) \tag{5.3}$$

of arbitrary maximal subgroups $h(\rho)$ of elements of g with rank ρ, and such groups $h(\rho)$ are maximal subgroups of elements of G with rank ρ.

It follows from the preceding lemma that the groups $h(\rho)$ admit a direct sum and that this direct sum g' is a group of elements with rank. For if $u_1, \cdots, u_m$ are non-null elements of different groups $h(\rho)$, Lemma 5.1 affirms

[12] See Hausdorff, l.c., p. 55.

replaced by comparisons involving isomorphisms of groups. A comparison by means of isomorphisms in general would not be very effective. Our comparisons will always be by means of isomorphisms of a special sort the abstract counterpart of which we now define.

Two elements of G will be said to be in the same *ρ-class* if they have the same rank while their difference has no rank or a lesser rank. An isomorphism between two subgroups of G of elements with rank will be termed a *ρ-isomorphism* if corresponding non-null elements are in the same ρ-class. These ρ-isomorphisms will be exemplified by three classes of *natural isomorphisms* defined by F.

We state the following theorem.

THEOREM 5.1. *Any two maximal groups of elements of G with a fixed rank σ are ρ-isomorphic.*

Let g_σ be the group generated by the elements of G with rank at most σ. Let H_σ be the subset of elements of g_σ without rank or with rank less than σ. We continue by proving the following statement.

(α). *The elements of H_σ form a group.*

Let u and v be arbitrary elements of H_σ. Writing E for exists and $\sim E$ for does not exist we distinguish four cases:

(1) $\rho(u)\ E,\ \rho(v)\ E.$ (2) $\rho(u)\ E,\ \rho(v) \sim E.$

(3) $\rho(u) \sim E,\ \rho(v)\ E.$ (4) $\rho(u) \sim E,\ \rho(v) \sim E.$

We shall prove that $u + v$ is in H_σ. If $\rho(u + v)$ does not exist $u + v$ is in H_σ. We assume therefore that $\rho(u + v)$ exists and seek to prove that $\rho(u + v) < \sigma$. This will follow from the rank conditions II, III, III, IV respectively in the cases (1), (2), (3), (4). In case (1), $\rho(u)$ and $\rho(v)$ are less than σ since u and v are in H_σ. It follows from II that $\rho(u + v) < \sigma$. In case (2), $\rho(u) < \sigma$ and $\rho(u + v)$ must be less than σ for otherwise $\rho(v)$ would exist. Case (3) is similar. The result in case (4) follows from IV, and the proof of (α) is complete.

We return to the theorem and let m_σ be a maximal group of elements of G with rank σ. Each element of m_σ is in a coset of the residue group h_σ defined by g_σ mod H_σ, and different elements u and v of m_σ are in different cosets of h_σ, for otherwise $u - v$ would be in H_σ and hence not in m_σ. Moreover there s an element of m_σ in each coset z of h_σ for otherwise an element u of z taken with m_σ would generate a group of elements of G with rank σ containing m_σ as a proper subgroup. There is thus an isomorphism between m_σ and h_σ in which each element u of m_σ corresponds to the coset of h_σ which contains u. We term m_σ and h_σ *ρ-isomorphic* since u corresponds to a coset of elements in the same ρ-class as u.

Let m'_σ be a second maximal group of elements of G with rank σ. The group m'_σ is also ρ-isomorphic with h_σ. It follows that m_σ and m'_σ are ρ-isomorphic with each other, and the proof of the theorem is complete.

A particular consequence of the theorem is that the dimensions of any two maximal groups of elements of G with rank σ are equal, but the theorem of course implies much more than this equality of dimension.

We note another consequence of statement (α).

(β). *The property of elements with rank being in the same ρ-class is transitive.* For if u, v, and w have the rank σ and if u is in the ρ-class of v and w, $u - v$ and $u - w$ belong to H_σ. It follows from (α) that $v - w$ belongs to H_σ so that v and w are in the same ρ-class.

We state the following lemma.

LEMMA 5.1. *If $u_1, \cdots, u_m$ are elements of G with ranks such that*

$$\rho(u_1) > \rho(u_i) \qquad (i = 2, \cdots, m) \tag{5.2}$$

then $\rho(u_1 + \cdots + u_m)$ exists and equals $\rho(u_1)$.

Suppose first that $m = 2$. By virtue of rank condition III $\rho(u_1 + u_2)$ exists. From I we see that $\rho(-u_2) = \rho(u_2)$. Since $u_1 = (u_1 + u_2 - u_2)$ we can infer from II and (5.2) that

$$\rho(u_1) \leqq \max [\rho(u_1 + u_2), \rho(u_2)] \leqq \rho(u_1 + u_2),$$

and that

$$\rho(u_1 + u_2) \leqq \max [\rho(u_1), \rho(u_2)] \leqq \rho(u_1)$$

so that $\rho(u_1) = \rho(u_1 + u_2)$. The proof of the lemma can be completed by induction with respect to m.

Let there be given a subgroup g of G and a set of subgroups $h(\alpha)$ of g, α being an enumerating index. The group g is said to be a *direct sum*

$$g = \sum_\alpha h(\alpha)$$

of the groups $h(\alpha)$ if each element u of g is a finite sum of elements from the groups $h(\alpha)$, and if there exists no relation of the form

$$u_{\alpha_1} + \cdots + u_{\alpha_m} = 0$$

in which the α_i's are distinct and u_{α_i} is a non-null element from the group $h(\alpha_i)$.

From this point on we shall frequently employ the hypothesis that the ranks ρ are well-ordered. We shall not mean thereby that the ranks *can* be well-ordered but that taken in their given orders they *are* well-ordered.[12]

We state the following theorem.

THEOREM 5.2. *If the ranks of elements of G are well-ordered a maximal subgroup g of the elements of G with rank is a direct sum*

$$\sum_\rho h(\rho) \tag{5.3}$$

of arbitrary maximal subgroups $h(\rho)$ of elements of g with rank ρ, and such groups $h(\rho)$ are maximal subgroups of elements of G with rank ρ.

It follows from the preceding lemma that the groups $h(\rho)$ admit a direct sum and that this direct sum g' is a group of elements with rank. For if $u_1, \cdots, u_m$ are non-null elements of different groups $h(\rho)$, Lemma 5.1 affirms

[12] See Hausdorff, l.c., p. 55.

that the rank of their sum v exists. Hence in particular $v \neq 0$. It remains to prove that $g' = g$.

To that end let u be an arbitrary non-null element of g. Set $\rho(u) = \sigma$. Since $h(\sigma)$ is a maximal subgroup of elements of g with rank σ there exists, in accordance with (5.1), an element v_1 in $h(\sigma)$ and an element w_1 not an element with rank σ, such that $u = v_1 + w_1$. It follows from the rank condition II that

Case I. Either $\rho(w_1)$ fails to exist,
Case II. Or $\rho(w_1) < \sigma$.

In Case I, $w_1 = 0$ since $w_1 = u - v_1$ and is in g. In Case II we treat $u - v_1$ as we originally treated u.

Proceeding inductively we infer the existence of a set of elements v_i, $i = 1, 2, \cdots, n$, respectively with decreasing ranks ρ_i, such that v_i is in $h(\rho_i)$ and

$$u = v_1 + \cdots + v_n + w_n$$

where either $\rho(w_n)$ fails to exist or $\rho(w_n) < \rho_n$. But a decreasing sequence of ranks is finite since the ranks are well-ordered, so that $\rho(w_n)$ fails to exist for some least integer n, say m. Since w_m is in g, $w_m = 0$ and $u = v_1 + \cdots + v_m$. Thus g equals the sum (5.3) as stated.

To prove the final statement of the theorem let $k(\rho)$ be any subgroup of elements of G with rank ρ, so chosen that $k(\rho) \supset h(\rho)$. It follows from Lemma 5.1 that the groups $k(\rho)$ have a direct sum J which is a group of elements with rank. Moreover $J \supset g$. Since g is maximal we conclude that $J = g$. It follows that $k(\rho) = h(\rho)$ since $h(\rho)$ is maximal in g, and we conclude that $h(\rho)$ is a maximal subgroup of elements of G with rank ρ.

The proof of the theorem is complete.

The following example shows that the omission of the hypothesis that the ranks are well-ordered in Theorem 5.2 would lead to a false conclusion.

Example 5.1. Let G be generated by a countable infinity of symbols $a_1, a_2, \cdots$ with coefficients which are integers mod 2. We suppose that a_n has a rank ρ_n and that $\rho_1 > \rho_2 > \cdots$ while the rank of a finite sum of different symbols a_i equals the maximum of the ranks of the summands. Every element of G except the null element thereby possesses a rank, and the four rank conditions are satisfied. The element $v_n = a_n + a_{n+1}$ taken with 0 forms a maximal subgroup (v_n) of elements of G with rank ρ_n. But the group G is not the direct sum $\sum_n (v_n)$ since the element a_1 is not in this direct sum.

Theorem 5.2 is complemented by the following theorem.

THEOREM 5.3. *If the ranks of elements of G are well-ordered, the direct sum of arbitrary maximal subgroups of elements of G with the respective ranks* ρ *is a maximal subgroup of elements of G with rank.*

Let $h(\rho)$ be a maximal subgroup of elements of G with rank ρ. As we have seen in the preceding proof the direct sum

$$H = \sum_{\rho} h(\rho)$$

is permissible and defines a group of elements with rank. We seek to prove that H is maximal in G.

To that end let g be a subgroup of elements of G with rank, with $g \supset H$. The group $h(\rho)$ is maximal in G and hence in g. It follows from the preceding theorem that

$$g = \sum_{\rho} h(\rho),$$

and the proof of the theorem is complete.

Note. The proofs of Theorems 5.2, 5.3 and 5.5 do not use rank condition IV. We shall prove the following theorem.

THEOREM 5.4. *If the ranks of elements of G are well-ordered any two maximal subgroups g and g' of elements of G with rank are ρ-isomorphic.*

By virtue of Theorem 5.2 g and g' are direct sums

$$g = \sum_{\rho} g(\rho), \qquad g' = \sum_{\rho} g'(\rho)$$

of subgroups $g(\rho)$ and $g'(\rho)$ of g and g' respectively, which are maximal subgroups of elements of G with rank ρ. According to Theorem 5.1 $g(\rho)$ and $g'(\rho)$ are ρ-isomorphic. These isomorphisms between the respective groups $g(\rho)$ and $g'(\rho)$ imply a unique isomorphism T between g and g' defined as follows. An arbitrary element u of the group g is of the form

$$u = u_1 + \cdots + u_m,$$

where the elements u_i are from different groups $g(\rho_i)$. If u_i' is the element of $g'(\rho_i)$ which corresponds to u_i, then the element u shall correspond under T to the element

$$u' = u_1' + \cdots + u_m'.$$

We shall show that T is a ρ-isomorphism. Without loss of generality we can suppose that

$$\rho_1 > \rho_2 > \cdots > \rho_m.$$

It follows from Lemma 5.1 that u and u' are respectively in the ρ-classes of u_1 and u_1'. But u_1 and u_1' are in the same ρ-class. Hence u and u' are in the same ρ-class since the property of elements being in the same ρ-class is transitive. The proof of the theorem is complete.

Extensions of the preceding theorems. If H is an arbitrary subgroup of G it is readily seen that the ranks ρ of elements of H by themselves satisfy the four rank conditions. The preceding theorems accordingly admit an immediate extension in which G is replaced by H. Thus regardless of whether the ranks of G are well-ordered or not as a whole, the preceding theorems hold upon replacing G by H, provided the ranks of H alone are well-ordered. The case where all the elements of H except the null element have rank will be met in the proof of the next theorem.

Bi-lexical ranks. Let s and t be variables which range over a simply ordered

set of elements (x). We suppose that the ranks $\rho(u)$ of elements u of G are given by lexicographically ordered pairs (s, t). If an element u of G has a rank $\rho = (s, t)$ we shall write $s = s(u)$ and $t = t(u)$ and term s and t the first and second components of ρ respectively. We shall assume that $\rho(u)$ exists if and only if $s(u)$ exists, and that $s(u)$ in turn exists if and only if $t(u)$ exists. Ranks $\rho = (s, t)$ satisfying these conditions will be termed *bi-lexical.*

The ranks $r(u)$ of F-non-bounding k-cycles are bi-lexical.

We shall prove the following theorem.

THEOREM 5.5. *When the ranks $\rho = (s, t)$ of elements u of G are bi-lexical and well-ordered each maximal group of elements of G with rank is a direct sum of maximal groups of elements of G with the respective values of t, and arbitrary maximal groups of elements of G with the respective values of t sum to a maximal group of elements of G with rank.*

A maximal group g of elements of G with rank is a direct sum of maximal groups $g(\rho)$ of elements of G with the respective ranks ρ, as we have seen in Theorem 5.3. Let $h(t)$ be the direct sum of those groups $g(\rho)$ for which the second component of ρ is t. Then g is a direct sum

$$g = \sum_t h(t). \tag{5.4}$$

Moreover each group $h(t')$ must be a maximal group of elements of G for which $t(u) = t'$. For otherwise there would exist an element v such that $t(v) = t'$ and such that the direct sum of the group (v) generated by v and $h(t')$, is a group of elements u for which $t(u) = t'$. If u is an element of the group,

$$H = \sum h(t) \qquad (t \neq t') \tag{5.5}$$

then $t(u) \neq t'$, so that the ranks of elements of $(v) + h(t')$ and H are different. Hence we can form the direct sum

$$K = (v) + h(t') + H,$$

obtaining thereby a group of elements with rank. Moreover K contains g as a proper subgroup. From this contradiction we infer that the group $h(t')$ is a maximal group of elements of G for which $t(u) = t'$, and the proof of the first affirmation in the theorem is complete.

We shall now prove the second affirmation in the theorem.

To that end let $k(t)$ be an arbitrary maximal subgroup of elements u of G for which $t(u) = t$. Let $g(s, t)$ be a maximal subgroup of elements v of $k(t)$ for which $s(v) = s$. By virtue of the extension of the preceding theorems noted just before the present theorem, $k(t)$ is a direct sum

$$k(t) = \sum_s g(s, t).$$

I say that $g(s, t)$ is a maximal group of elements u of G for which $\rho = (s, t)$. Otherwise we could reason as in the proof that $h(t)$ was maximal and obtain a group $p(t)$ of elements v for which $t(v) = t$ and which would contain $k(t)$ as a

proper subgroup. Hence $g(s, t)$ is a maximal group of elements of G with rank $\rho = (s, t)$. It follows from Theorem 5.3 that the group

$$\sum_{t,s} g(s, t) = \sum_t k(t)$$

is a maximal group of elements of G with rank, and the proof of the theorem is complete.

$*$-Telescopic group sequences. Our remarks in the remainder of this section are largely by way of orientation.

In earlier papers on critical points of functions the author has been concerned with integers $\mu_k = M_k - R_k$, positive or zero, where M_k was the k^{th} type number sum of the critical sets and R_k was the k^{th} connectivity of M. The numbers μ_k, when finite, were found to be of the form

$$\mu_k = \gamma_k + \gamma_{k-1} \qquad (\gamma_k \geqq 0; k = 0, 1, \cdots) \tag{5.6}$$

where γ_k is an integer and $\gamma_{-1} = 0$.

We continue with the following lemma.

LEMMA 5.2. *A necessary and sufficient condition that a sequence of non-negative integers $\mu_0, \mu_1, \cdots$ be representable in the form (5.6) is that these integers satisfy the inequalities*

$$\mu_n - \mu_{n-1} + \cdots + (-1)^n \mu_0 \geqq 0 \qquad (n = 0, 1, \cdots). \tag{5.7}$$

If the numbers μ_i are representable in the form (5.6) we find that

$$\mu_n - \mu_{n-1} + \cdots + (-1)^n \mu_0 = \gamma_n \tag{5.8}$$

so that the inequalities (5.7) are a consequence of (5.6). Conversely if the relations (5.7) are satisfied and we define γ_k by means of (5.8), we find that $\gamma_k \geqq 0$ and that (5.6) holds.

The inequalities (5.7) were fundamental in the earlier theory and were the source of numerous existence theorems for critical points.

The present more general theory is formally analogous to the earlier theory with groups m_k replacing the numbers μ_k and the group operation of reduction with respect to a modulus replacing the operation of subtraction in (5.7). In this connection it should be noted that (5.7) can be written in the form

$$\mu_n - [\mu_{n-1} - (\mu_{n-2} \cdots \mu_0) \cdots] \geqq 0, \tag{5.9}$$

employing subtractions only.

More explicitly we shall be concerned with an infinite sequence of groups $m_0, m_1, \cdots$. The function F will give rise to certain natural isomorphisms termed $*$-isomorphisms, which map a group g_k onto a group denoted by $\overset{*}{g}_k$. In terms of these $*$-isomorphisms a sequence of groups m_k will be termed *$*$-telescopic* if there exists an auxiliary sequence of groups g_k such that

$$m_k = g_k + \overset{*}{g}_{k-1} \qquad (g_{-1} = 0; k = 0, 1, \cdots). \tag{5.10}$$

The relations (5.10) are the formal analogues of the relations (5.6). In fact if μ_k and γ_k are the dimensions of m_k and g_k respectively the relations (5.6) are a consequence of the relations (5.10).

We can also give a formal analogue of the relations (5.9). To that end we read mod* A, "mod a group $*$-isomorphic with A." We then replace (5.9) by the relations

$$(5.11)\quad m_n \text{ mod* } (m_{n-1} \text{ mod* } (m_{n-2} \cdots \text{ mod* } (m_0) \cdots) \geqq 0, \quad (n = 0, 1, \cdots)$$

reading $\geqq 0$ "exists." In (5.11) as in (5.9) the operations indicated are to be successively performed starting with the innermost parenthesis. The symbolic expressions

$$(5.12)\quad \begin{array}{c} m_1 \text{ mod* } m_0 , \\ m_2 \text{ mod* } m_1 \text{ mod* } m_0 , \\ \cdots\cdots\cdots\cdots\cdots\cdots , \end{array}$$

appear in infinitely many of the relations (5.11), but in each of these relations they shall indicate the same groups.

It is easily seen that a necessary and sufficient condition that the groups m_k satisfy relations of the form (5.11) is that there exist auxiliary groups g_k such that the groups m_k are of the form (5.10) for suitably defined $*$-isomorphisms These $*$-isomorphisms will presently receive a concrete exemplification.

II. Natural Isomorphisms

6. **k-caps and a-isomorphisms.** The object of chapter II is to establish certain general relations between groups of F-non-bounding i-cycles and groups of j-caps. We shall make use of three basic types of isomorphisms, r-isomorphisms between groups of F-non-bounding k-cycles, a-isomorphisms between groups of k-caps and $*$-isomorphisms between groups of F-non-bounding k-cycles and groups of (k + 1-caps. These isomorphisms depend in an essential way upon the nature of F.

We first introduce several new concepts.

k-caps. A k-cycle u on F_{a+} mod_o F_a will be termed a k-cap *belonging* to a if

$$u \not\sim 0 \qquad (\text{on } F_{a+} \text{ mod}_o\, F_a).$$

The constant a will then be termed a *k-cap limit* $a(u)$ of u. The k-cap limits $a(u)$ satisfy the four rank axioms $a(u)$ replacing $\rho(u)$, provided the underlying group G of the preceding section be correctly defined.

To define G we term a sequence w of algebraic k-chains w_n with norms e_n tending to zero a *formal* k-chain. To add two formal k-chains we add their components, while δw shall be the formal k-chain with components δw_n. Formal k-chains make up an additive abelian operator group with operators δ. The components of a formal k-chain w may in particular be the components of a k-cap z. In such a case it is easily seen that $a(z)$ is uniquely determined by w.

To add two k-caps u and v consider the formal k-chain $w = u + v$. If w defines a k-cap z we regard z as the cap sum of u and v. If w defines no k-cap we understand that there is no cap sum $u + v$. In the group of formal k-chains G certain formal k-chains are thus k-caps and possess cap limits. Concerning those cap limits we have the following theorem.

THEOREM 6.1. *The k-cap limits $a(u)$ satisfy the rank conditions, $a(u)$ replacing $\rho(u)$ with the underlying group G the group of formal k-chains.*

The proof of this theorem is immediate.

We shall give certain definitions which concern k-caps. We shall term an homology

$$u \sim 0 \qquad (\text{on}_o\ F_{c+}\ \text{mod}_o\ F_c)$$

a *c-homology*. The earlier ρ-class and ρ-isomorphism are here replaced by their formal analogues *a-class* and *a-isomorphism*. It is seen that two k-caps with cap limits c belong to the same a-class if and only if their difference is c-homologous to zero.

Let p be a point on M and c a number such that $0 \leqq c \leqq 1$. We term (p, c) a *limiting pair* if c is a limit at p of values of $F(p_n)$ for some sequence p_n tending to p. We include pairs of the form $[p, F(p)]$ as limiting pairs. There are certain limiting pairs which are particularly distinguished and which will be called critical or semi-critical pairs. They will be defined in §9. Their existence depends upon a series of theorems which we shall now develop.

A k-cap which is a k-cycle will be termed a *k-cycle-cap*. The following theorem may be regarded as a first existence theorem in the theory of k-caps.

THEOREM 6.2. *Corresponding to an arbitrary F-non-bounding k-cycle u there is a k-cycle-cap z in the r-class of u such that the cap limit $a(z)$ differs arbitrarily little from the inferior cycle limit $t(u)$.*

If $t(u) = 1$, u is a cycle-cap and the theorem is true. We suppose then that $t(u) < 1$.

If the r-class of u contains a cycle-cap belonging to $t(u)$ the theorem is again true. If the r-class of u contains no such cycle-cap let β be an arbitrary constant less than 1 such that $t(u) < \beta < s(u)$.

There exists a k-cycle u' in the given r-class below$_o$ β with $u' \nsim 0$ on $F \leqq \beta$. Let α be the inferior cycle limit of u' on $F \leqq \beta$. It follows from Lemma 2.2 that $u' \sim 0 \bmod F \leqq \alpha$ on $F \leqq \beta$. Upon setting $F_{\alpha+} = V$ and $F_{\beta+} = C$ with $U = 0$ it follows from Theorem 3.1 that there exists a k-cycle z on V such that $u' \sim z$ on C. But z is not α-homologous to zero for otherwise α would not be an inferior cycle limit of z on $F \leqq \beta$. Hence z is a cycle-cap belonging to α. The theorem follows directly.

An F-non-bounding k-cycle u which lies on the domain $F \leqq t(u)$ will be termed *canonical*. We have seen that there is a cycle-cap in the r-class of each F-non-bounding k-cycle u, but there need be no canonical k-cycle in the r-class of u as the following example shows.

The relations (5.10) are the formal analogues of the relations (5.6). In fact if μ_k and γ_k are the dimensions of m_k and g_k respectively the relations (5.6) are a consequence of the relations (5.10).

We can also give a formal analogue of the relations (5.9). To that end we read mod* A, "mod a group $*$-isomorphic with A." We then replace (5.9) by the relations

$$(5.11) \quad m_n \operatorname{mod}^* (m_{n-1} \operatorname{mod}^* (m_{n-2} \cdots \operatorname{mod}^* (m_0) \cdots) \geqq 0, \quad (n = 0, 1, \cdots)$$

reading $\geqq 0$ "exists." In (5.11) as in (5.9) the operations indicated are to be successively performed starting with the innermost parenthesis. The symbolic expressions

$$(5.12) \quad \begin{array}{l} m_1 \operatorname{mod}^* m_0 , \\ m_2 \operatorname{mod}^* m_1 \operatorname{mod}^* m_0 , \\ \dots\dots\dots\dots\dots\dots , \end{array}$$

appear in infinitely many of the relations (5.11), but in each of these relations they shall indicate the same groups.

It is easily seen that a necessary and sufficient condition that the groups m_k satisfy relations of the form (5.11) is that there exist auxiliary groups g_k such that the groups m_k are of the form (5.10) for suitably defined $*$-isomorphisms These $*$-isomorphisms will presently receive a concrete exemplification.

II. Natural Isomorphisms

6. ***k*-caps and *a*-isomorphisms.** The object of chapter II is to establish certain general relations between groups of F-non-bounding i-cycles and groups of j-caps. We shall make use of three basic types of isomorphisms, r-isomorphisms between groups of F-non-bounding k-cycles, a-isomorphisms between groups of k-caps and $*$-isomorphisms between groups of F-non-bounding k-cycles and groups of $(k + 1$-caps. These isomorphisms depend in an essential way upon the nature of F.

We first introduce several new concepts.

k-caps. A k-cycle u on F_{a+} mod$_o$ F_a will be termed a k-cap *belonging* to a if

$$u \not\sim 0 \qquad (\text{on } F_{a+} \operatorname{mod}_o F_a).$$

The constant a will then be termed a *k-cap limit* $a(u)$ of u. The k-cap limits $a(u)$ satisfy the four rank axioms $a(u)$ replacing $\rho(u)$, provided the underlying group G of the preceding section be correctly defined.

To define G we term a sequence w of algebraic k-chains w_n with norms e_n tending to zero a *formal* k-chain. To add two formal k-chains we add their components, while δw shall be the formal k-chain with components δw_n. Formal k-chains make up an additive abelian operator group with operators δ. The components of a formal k-chain w may in particular be the components of a k-cap z. In such a case it is easily seen that $a(z)$ is uniquely determined by w.

To add two k-caps u and v consider the formal k-chain $w = u + v$. If w defines a k-cap z we regard z as the cap sum of u and v. If w defines no k-cap we understand that there is no cap sum $u + v$. In the group of formal k-chains G certain formal k-chains are thus k-caps and possess cap limits. Concerning those cap limits we have the following theorem.

THEOREM 6.1. *The k-cap limits $a(u)$ satisfy the rank conditions, $a(u)$ replacing $\rho(u)$ with the underlying group G the group of formal k-chains.*

The proof of this theorem is immediate.

We shall give certain definitions which concern k-caps. We shall term an homology

$$u \sim 0 \qquad (\text{on}_\circ\ F_{c+} \ \text{mod}_\circ\ F_c)$$

a *c-homology*. The earlier ρ-class and ρ-isomorphism are here replaced by their formal analogues *a-class* and *a-isomorphism*. It is seen that two k-caps with cap limits c belong to the same a-class if and only if their difference is c-homologous to zero.

Let p be a point on M and c a number such that $0 \leqq c \leqq 1$. We term (p, c) a *limiting pair* if c is a limit at p of values of $F(p_n)$ for some sequence p_n tending to p. We include pairs of the form $[p, F(p)]$ as limiting pairs. There are certain limiting pairs which are particularly distinguished and which will be called critical or semi-critical pairs. They will be defined in §9. Their existence depends upon a series of theorems which we shall now develop.

A k-cap which is a k-cycle will be termed a *k-cycle-cap*. The following theorem may be regarded as a first existence theorem in the theory of k-caps.

THEOREM 6.2. *Corresponding to an arbitrary F-non-bounding k-cycle u there is a k-cycle-cap z in the r-class of u such that the cap limit $a(z)$ differs arbitrarily little from the inferior cycle limit $t(u)$.*

If $t(u) = 1$, u is a cycle-cap and the theorem is true. We suppose then that $t(u) < 1$.

If the r-class of u contains a cycle-cap belonging to $t(u)$ the theorem is again true. If the r-class of u contains no such cycle-cap let β be an arbitrary constant less than 1 such that $t(u) < \beta < s(u)$.

There exists a k-cycle u' in the given r-class below$_\circ$ β with $u' \nsim 0$ on $F \leqq \beta$. Let α be the inferior cycle limit of u' on $F \leqq \beta$. It follows from Lemma 2.2 that $u' \sim 0 \bmod F \leqq \alpha$ on $F \leqq \beta$. Upon setting $F_{\alpha+} = V$ and $F_{\beta+} = C$ with $U = 0$ it follows from Theorem 3.1 that there exists a k-cycle z on V such that $u' \sim z$ on C. But z is not α-homologous to zero for otherwise α would not be an inferior cycle limit of z on $F \leqq \beta$. Hence z is a cycle-cap belonging to α. The theorem follows directly.

An F-non-bounding k-cycle u which lies on the domain $F \leqq t(u)$ will be termed *canonical*. We have seen that there is a cycle-cap in the r-class of each F-non-bounding k-cycle u, but there need be no canonical k-cycle in the r-class of u as the following example shows.

Example 6.1. Let the space M consist of the closure of the curve

$$y = (1 - x) \sin \frac{1}{x} \qquad (0 < x \leqq 1)$$

together with a straight line segment joining the point $(0, 1)$ to the point $(-1, -1)$ in the (x, y) plane. On M let $F = (y + 1)/2$ so that $0 \leqq F \leqq 1$. Let u be a 0-cycle mod 2 whose components u consist of the two points $(-1, -1)$ and $(1, 0)$. We see that $t(u) = 0$. There is no 0-cycle on the domain $F = 0$ in the r-class of u.

Example 6.2. We can modify the preceding example to show that an inferior cycle limit may equal no cap limit. To that end we join the bottom points $(-1, -1)$, $(-1, 0)$ of the preceding point set M by a semi-circle on the domain $y \leqq -1$ and let the resulting point set be denoted by M'. On M' let $F = y$ waiving the conditions that $0 \leqq F \leqq 1$. The 0-cycle u of the preceding example now has an inferior cycle limit which equals no cap limit, since the semi-circle can be deformed on itself away from its end points.

7. *-Isomorphisms. If v is a k-cap such that $\beta v \sim 0$ below$_\circ$ its cap limit $a(v)$, v will be termed *linkable*. A non-linkable k-cap v will be said to *cap* its boundary $u = \beta v$ and u will then be said to admit the cap v. A group of F-non-bounding k-cycles will be said to be **-isomorphic* with a group g_k^* of $(k + 1)$-caps if g_k and g_k^* admit an isomorphism in which each k-cycle of g_k is capped by its correspondent in g_k^*.

We shall make use of the following lemma.

LEMMA 7.1. *If u is a k-cycle mod$_\circ$ F_a on $F \leqq a$ and is a-homologous to zero, then $\beta u \sim 0$ below$_\circ$ a.*

Under the hypotheses of the lemma we can affirm the existence of a positive constant e with the following property. Corresponding to the n[th] component u_n of u there exists an algebraic k-chain w_n on $F \leqq a - e$ and an algebraic $(k + 1)$-chain z_n on $F \leqq a$ such that $\beta z_n = u_n + w_n$ where u_n, w_n, and z_n possess norms e_n which tend to zero as n becomes infinite. Hence $\beta u_n + \beta w_n = 0$ and $\beta u_n \sim_{e_n} 0$ on $F \leqq a - e$.

The proof of the lemma is complete.

An F-non-bounding k-cycle u which admits a k-cap v is bounding. A k-cycle which is both F-non-bounding and bounding has been termed ambiguous. With this understood we state the following theorem.

THEOREM 7.1. *An ambiguous k-cycle u admits at least one $(k + 1)$-cap v, with $a(v) = s(u)$.*

Let $s(u)$ be denoted by a. If e is an arbitrary positive constant then $u \sim 0$ below$_\circ$ $a + e$ by virtue of the definition of $s(u)$. If $a < 1$ it follows from Lemma 2.2 that $u \sim 0$ on $F \leqq a$. This homology is trivial if $a = 1$.

We shall apply Theorem 3.2. To that end we set $A = 0$, and let V be a carrier of the homology $u \sim 0$ on $F \leqq a$. The cycle u has a carrier U on V below$_\circ$ a. It follows from Theorem 3.2 that there exists a $(k + 1)$-cycle

v mod U on V such that $\beta v = u$. Moreover v is not a-homologous to zero for otherwise $\beta v \sim 0$ below$_o$ a by virtue of Lemma 7.1, and we could infer that $u \sim 0$ below$_o$ a contrary to the nature of $a = s(u)$.

Thus v caps u, and the proof is complete.

Let g_k be a group of ambiguous k-cycles representable as a direct sum

$$g_k = \sum_c g_k(c) \tag{7.1}$$

of subgroups $g_k(c)$ of k-cycles u for which $s(u) = c$, where c ranges over an arbitrary set of superior cycle limits s. We term g_k *s-neuclear*, and state the following theorem.

THEOREM 7.2. *An s-neuclear group g_k of ambiguous k-cycles is $*$-isomorphic with at least one group g_k^* of $(k + 1)$-caps.*

Let $(z)_c$ be a maximal linear set in $g_k(c)$ and let (z) be a set theoretic sum of the sets $(z)_c$. The set (z) will form a maximal linear set for g_k. With each cycle z_i in (z) we now associate a $(k + 1)$-cap $K(z_i)$, as is possible by virtue of Theorem 7.1. We set $K(0) = 0$. Let $z_1, \cdots, z_n$ be a set of distinct cycles in (z). If $u = \delta_i z_i$ we set

$$K(u) = \delta_i K(z_i),$$

and prove statements (α) and (β).

(α). *The cycle u is capped by $K(u)$ if $u \neq 0$.*

In proving (α) we can assume without loss of generality that no δ_i is null. Set $a = s(u)$. Recalling that $a = \max s(z_i)$ we see that each $K(z_i)$ lies on $F \leqq a$, and hence $K(u)$ lies on this same domain. Moreover

$$\beta K(u) \doteq \delta_i z_i = u.$$

Finally $K(u)$ is not a-homologous to zero, for otherwise $u \sim 0$ below$_o$ a by virtue of Lemma 7.1.

The proof of (α) is complete.

(β). *The $(k + 1)$-caps $K(u)$ form a group g_k^* $*$-isomorphic with the group g_k.*

Let x and y be two cycles of g_k of the form

$$x = \delta_i z_i, \qquad y = \delta_i' z_i,$$

where the z_i's form a set of distinct cycles of (z). We see that

$$K(x + y) = \Sigma(\delta_i + \delta_i')\, K(z_i) = K(x) + K(y).$$

Hence the caps $K(u)$ define a group homeomorphic with g_k. That this homeomorphism is simple follows from the fact that $K(u) = 0$ only if $u = 0$, in accordance with (α).

The proof of the theorem is complete.

8. **Well-ordered cap limits.** The hypothesis that the cap limits are well-ordered has important consequences which will be studied in this section.

THEOREM 8.1. *Let u be an F-non-bounding k-cycle. If the k-cap limits*

greater than $t(u)$ are bounded away from $t(u)$ there is a canonical k-cycle in the r-class of u.

By virtue of Theorem 6.2 there exists a k-cycle-cap z in the r-class of u such that $a(z)$ differs arbitrarily little from $t(u)$. Under the hypotheses of the present theorem $a(z) = t(u)$ if $a(z) - t(u)$ is sufficiently small. But if $a(z) = t(u)$, z will be a canonical k-cycle in the r-class of u and the proof of the theorem is complete.

COROLLARY. *If the k- and $(k + 1)$-cap limits are well-ordered each k-cycle limit is a k- or $(k + 1)$-cap limit and the ranks of F-non-bounding k-cycles are well-ordered.*

That each superior cycle limit of a k-cycle is a $(k + 1)$-cap limit follows from Theorem 7.1, while inferior cycle limits of k-cycles are k-cap limits by virtue of the preceding theorem.

The ranks $r = (s, t)$ of F-non-bounding k-cycles can be grouped in classes with common s, and these classes will be well-ordered among themselves by virtue of the well-ordering of the $(k + 1)$-cap limits s. The ranks r in each such class can now be further ordered according to the values of t. Each class will then be well-ordered. The ranks thus receive their prescribed lexicographic order and it is readily seen that they are thereby well-ordered.

Let u be a k-cycle $\text{mod}_\circ\, H$ where H is a domain F_a or F_{a+}. A constant $b > a$ will be called a *homology limit* of u $\text{mod}_\circ\, H$ if u is $\text{on}_\circ\, F_b$ and $u \not\sim 0$ $\text{below}_\circ$ b $\text{mod}_\circ\, H$. A k-cap u with cap limit a will be termed *degenerate* if u possesses no homology limit $\text{mod}_\circ\, F_a$. That degenerate k-caps exist is shown by the following example.

Example 8.1. Let the space M consist of the points (x, y) on the curve $y = x \sin 1/x$ for $0 \leqq x \leqq 1$ and let $F = (y + 1)/2$ on M. Let u_n be an algebraic 0-cycle mod 2 which coincides with the point $(0, 0)$. The Vietoris 0-cycle $u = (u_n)$ is a degenerate 0-cap.

We continue with the following theorem.

THEOREM 8.2. *Let u be a k-cap with cap limit a. If the $(k + 1)$-cap limits greater than a are bounded away from a, u is non-degenerate.*

The theorem is trivial when $a = 1$.

Suppose $a < 1$. Let b be a lower bound between 1 and a of the $(k + 1)$-cap limits greater than a. We shall prove the following statement.

(i). *The constant b is a homology limit of u $\text{mod}_\circ\, F_a$.*

Suppose (i) false. Then for some constant c between a and b and some constant $\alpha < a$, $u \sim 0$ on $F \leqq c$ mod $F \leqq \alpha$. Upon setting

$$A = F_{\alpha+}, \qquad U = F_{a+}, \qquad V = F_{c+}, \tag{8.1}$$

it follows from Theorem 3.2 that there exists a $(k + 1)$-cycle v mod U on V such that $\beta v \equiv u$ mod A. Moreover

$$v \not\sim 0 \qquad (\text{mod } F_{a+} \text{ on } F_{c+}), \tag{8.2}$$

for otherwise $\beta v \sim 0$ on U, and hence $u \sim 0$ mod A on U contrary to the nature of u as a k-cap.

From (8.2) it follows that there is a greatest lower bound t of numbers $e > a$ such that $v \sim 0 \bmod F \leqq e$ on $F \leqq c$. Upon referring to Lemma 2.2, setting $B = F_{t+}$ and $C = F_{e+}$ we see that

$$v \sim 0 \qquad (\text{mod } F_{t+} \text{ on } F_{e+}). \tag{8.3}$$

It follows from (8.2) and (8.3) that $t > a$. We next apply Theorem 3.1, setting

$$U = F_{a+}, \qquad V = F_{t+}, \qquad C = F_{c+},$$

and infer from (8.3) that there exists a $(k + 1)$-cycle $z \bmod F \leqq a$ on $F \leqq t$ homologous to $v \bmod F \leqq a$ on $F \leqq c$. It follows from the choice of t that v and hence $z \not\sim 0 \text{ mod}_o\, F_t$ on F_{c+}. Since z is on $F \leqq t$, t is a cap limit of z with $a < t < b$, contrary to the choice of b.

The proof of (i), and of the theorem is complete.

We shall prove the following theorem.

THEOREM 8.3. (1). *There is a k-cycle-cap v in the a-class of each linkable k-cap u.* (2). *If u is non-degenerate v is F-non-bounding.* (3). *If the k- and $(k + 1)$-cap limits are well-ordered v may be chosen so as to be canonical.*

We suppose that a is the cap limit of u and let B be a carrier of u on $F \leqq a$. Since u is linkable there exists a constant $\alpha < a$ such that $\beta u \sim 0$ on $F \leqq \alpha$. If we denote the latter domain by A, set $V = A + B$, and apply Theorem 3.3, statement (1) of Theorem 8.3 follows directly.

If u is non-degenerate v is likewise. The cycle v thus possesses a homology limit $\text{mod}_o\, F_a$ and a fortiori a homology limit so that (2) is true.

We now turn to the proof of (3).

If the $(k + 1)$-cap limits are well-ordered, the $(k + 1)$-cap limits greater than a are in particular bounded away from a so that u is non-degenerate in accordance with Theorem 8.2. As stated in (2), v is then F-non-bounding. If the k- and $(k + 1)$-cap limits are well-ordered, the ranks of F-non-bounding k-cycles are well-ordered, and by virtue of Theorem 8.1 there is a canonical k-cycle in each r-class of k-cycles. A maximal group $g(r)$ of k-cycles with rank r can accordingly be made up of canonical k-cycles. If $h(t)$ is the sum of the groups $g(r)$ for which the second component of r is t, $\sum_t h(t)$ will be a maximal group of k-cycles with rank. Each non-null cycle z in a group $h(t)$ will lie on $F \leqq t$ and have t as an inferior cycle limit. Hence each such cycle z will be canonical.

There accordingly exists a set $v_1, \cdots, v_n$ of canonical k-cycles from different groups $h(t)$ such that the k-cycle $w = v - (v_1 + \cdots + v_n)$ is without rank. The cycle w is not a k-cap, for otherwise w would be a non-degenerate k-cap and hence have rank. Hence v is in the a-class of $(v_1 + \cdots + v_n) = z$. Without loss of generality we can suppose that $t(v_1) > \cdots > t(v_n)$. The cycle-cap z is not necessarily canonical, but it is in the a-class of v_1, and v_1 is canonical. Thus u is in the a-class of the canonical cycle v_1, and the proof of the theorem is complete.

An a-class of k-caps will be said to be of *excess* type if it contains no cycle

or contains an ambiguous k-cycle. Otherwise the a-class will contain a non-bounding k-cycle but no ambiguous k-cycle and be said to be of *space* type. A k-cap will be said to be of excess or space type according as it is in an a-class of excess or space type. A linkable k-cap may be either an excess or space cap. A non-linkable k-cap is always an excess cap. An excess k-cap and a space k-cap are never in the same a-class, and accordingly their difference is always a k-cap.

We shall now prove four theorems on the composition of maximal groups of k-caps.

THEOREM 8.4. *If the k-cap limits are well-ordered, a maximal group M_k of k-caps is a direct sum $M_k = M_k' + M_k''$ of arbitrary maximal subgroups respectively of excess and space k-caps of M_k. The groups M_k' and M_k'' are maximal among k-caps in general. Conversely arbitrary maximal groups of excess and space k-caps sum to a maximal group of k-caps.*

We shall begin by proving the following statement.

(α). *A maximal group μ_k of k-caps with a cap limit c is a direct sum*

$$\mu_k = \mu_k' + \mu_k''$$

of maximal groups μ_k' and μ_k'' of excess and space k-caps respectively with cap limit c.

The subset of linkable k-caps of μ_k together with the null element form a group a_k so that μ_k is a direct sum $\mu_k = a_k + c_k$ in which c_k is a maximal subgroup of non-linkable k-caps of μ_k. The subset of excess k-caps of a_k together with the null element form a group α_k so that a_k is a direct sum $\alpha_k + \beta_k$ in which β_k is a maximal subgroup of space k-caps of a_k. Hence μ_k is a direct sum $\alpha_k + \beta_k + c_k$. We set $\mu_k' = \alpha_k + c_k$ and $\mu_k'' = \beta_k$. That μ_k' is a group of excess k-caps with cap limit c follows from the fact that a non-linkable and a linkable k-cap with cap limits c sum to a non-linkable k-cap with cap limit c. That μ_k' and μ_k'' are respectively maximal groups of excess and space k-caps with cap limit c follows from the relation $\mu_k = \mu_k' + \mu_k''$, the fact that μ_k is maximal, and the fact that an excess and a space k-cap with cap limit c sum to a k-cap with cap limit c.

The proof of (α) is complete.

We can now establish the first two statements in the theorem.

Since the cap limits satisfy the rank conditions it follows from Theorem 5.2 that a maximal group M_k of k-caps is a direct sum of groups $\mu_k(c)$ which are maximal groups of k-caps with the respective cap limits c. Each such group $\mu_k(c)$ is a sum $\mu_k'(c) + \mu_k''(c)$ of the nature described in (α). Upon setting

$$m_k' = \sum_c \mu_k'(c), \qquad m_k'' = \sum_c \mu_k''(c)$$

we see that $M_k = m_k' + m_k''$. The groups m_k' and m_k'' are respectively groups of excess and space k-caps. That they are maximal follows from the relation $M_k = m_k' + m_k''$ and the fact that an excess and a space k-cap always sum to a k-cap.

The proof of the first two statements in the theorems is complete.

To prove the final statement in the theorem we assume that M'_k and M''_k are respectively maximal groups of excess and space k-caps. The sum

$$L_k = M'_k + M''_k \tag{8.4}$$

is composed of k-caps. Let M_k be a maximal group of k-caps such that $M_k \supset L_k$. The groups M'_k and M''_k are maximal among formal k-chains and hence in M_k. It follows from the first statement in our theorem that

$$M_k = M'_k + M''_k.$$

The proof of the theorem is complete.

THEOREM 8.5. *If the k-cap limits are well-ordered for each k,*

(i) *there exists a maximal group g of ambiguous k-cycle-caps which is a maximal group of ambiguous k-cycles,*

(ii) *such a group g_k admits a $*$-isomorphic group of $(k + 1)$-caps $\overset{*}{g}_k$,*

(iii) *while the groups g_k and $\overset{*}{g}_{k-1}$ admit a direct sum*

$$m_k = g_k + \overset{*}{g}_{k-1} \qquad (k = 0, 1, \cdots) \tag{8.5}$$

which is a maximal group of excess k-caps.

We shall prove statement (i) with the aid of Theorem 5.5.

To that end we shall introduce a new set of ranks σ, assigning ranks σ only to ambiguous k-cycles u and for such cycles setting $\sigma(u) = r(u)$. The ranks σ clearly satisfy rank conditions I, II, III.

There is a canonical k-cycle in each r-class by virtue of Theorem 8.1. It follows that there exists a maximal group $k(t)$ of ambiguous k-cycles with a prescribed inferior cycle limit t with the property that the non-null cycles of $k(t)$ are canonical. We now apply Theorem 5.5 employing ranks σ. It follows that $\Sigma\, k(t)$ is a maximal group g of ambiguous k-cycles. But each cycle of $k(t)$ is a cycle-cap with cap limit t. The group g is thus a group of cycle-caps. Since g is maximal as a group of ambiguous k-cycles it is a fortiori maximal as a group of ambiguous k-cycle-caps.

The proof of statement (i) is complete.

Statement (ii) is a consequence of Theorem 7.2.

We continue with a proof of the following statement.

(α). *The groups g_k and $\overset{*}{g}_{k-1}$ admit a direct sum m_k which is a group of excess k-caps.*

Let u and v be k-caps in g_k and $\overset{*}{g}_{k-1}$ respectively. We seek to prove that $u + v$ is an excess k-cap. If $a(u) > a(v)$, $u + v$ is a k-cap in the a-class of u, and since u is of excess type $u + v$ is of excess type. The case where $a(v) > a(u)$ is similar. If $a(u) = a(v)$, $u + v$ is non-linkable and hence of excess type. Thus (α) is true.

(β). *If u is an arbitrary non-linkable k-cap there is an element w in $\overset{*}{g}_{k-1}$ such that $u - w$ is not a non-linkable k-cap with cap limit $a(u)$.*

Let $\beta u = v$. The cycle v is ambiguous so that there exists a cycle z in g_{k-1}

such that $v - z$ has no rank. Let w be the image of z in $\overset{*}{g}_{k-1}$. Since $v - z$ has no rank $v - z \sim 0$ below$_\circ$ $a(u)$. Hence

$$\beta(u - w) = v - z \sim 0 \qquad [\text{below}_\circ\ a(u)]. \tag{8.6}$$

If $a(u - w) = a(u)$, $u - w$ is linkable by virtue of (8.6) and (β) is true. If $a(u - w)$ does not exist or is less than $a(u)$ we see that (β) is again true.

(γ). *If u is an excess linkable k-cap there is an element w in g_k such that $u - w$ is not a k-cap with cap limit $a(u)$.*

Let v be an ambiguous k-cycle in the a-class of u. There exists a k-cycle w in g_k such that $v - w$ is not an ambiguous k-cycle and hence not a k-cap. Thus v, w, and u are in the same a-class. Since u and w are in the same a-class, $a(u - w)$ does not exist or is less than $a(u)$, and (γ) is true.

(δ). *If u is an excess k-cap there is an element w in m_k such that $u - w$ is not an excess k-cap with cap limit $a(u)$.*

In case u is linkable (δ) follows from (γ).

We accordingly assume that u is non-linkable. According to (β) there exists a k-cap w in $\overset{*}{g}_{k-1}$ such that $u - w$ is not a non-linkable k-cap with cap limit $a(u)$. The difference $u - w$ may satisfy (δ). Otherwise $a(u - w) = a(u)$ and $u - w$ is an excess linkable k-cap to which (γ) applies. Hence (δ) holds in all cases.

We can now prove (iii). That m_k is a group of excess k-caps has been proved in (α). It remains to show that m_k is maximal. To that end let u be an arbitrary excess k-cap and w the corresponding k-cap in (δ). If $u - w$ is not an excess k-cap (iii) is true. If $u - w$ is an excess k-cap $a(u - w) < a(u)$ in accordance with (δ). Set $w = w_1$. We now treat $u - w_1$ as we did u. Proceeding inductively and recalling that a decreasing sequence of k-cap limits is finite we conclude that there exists a set $w_1, w_2, \cdots w_m$ of k-caps in m_k with decreasing cap limits such that $u - (w_1 + \cdots + w_m)$ is not an excess k-cap.

The proof of (iii) and of the theorem is complete.

Upon referring to Lemma 5.2 we have the following corollary.

COROLLARY. *The dimension μ_k of the maximal groups m_k of excess k-caps, if finite, satisfy the relations* (5.7). *More generally the groups m_k are $*$-telescopic and satisfy relations of the form* (5.11).

We come finally to maximal groups of space k-caps. We shall prove the following theorem.

THEOREM 8.6. *If the k- and $(k + 1)$-cap limits are well-ordered each maximal group of canonical non-bounding k-cycles is*

(1) *a maximal group of non-bounding k-cycles,*

(2) *a maximal group of space k-caps.*

We shall first prove statement (1) of the theorem.

To that end we assign ranks σ to canonical non-bounding k-cycles setting $\sigma(u) = r(u)$. For such cycles $s(u) = +\infty$. The rank conditions I, II, III are satisfied by ranks σ.

Let $g(\sigma)$ be a maximal group of canonical non-bounding k-cycles with rank σ. It follows from Theorems 5.2 and 5.3 that the sum

$$g = \sum_{\sigma} g(\sigma) \tag{8.7}$$

is a maximal group of canonical non-bounding k-cycles and each maximal group of canonical non-bounding k-cycles is of the form (8.7). But there is a canonical k-cycle in each r-class of k-cycles. It follows that $g(\sigma)$ is a maximal group of non-bounding k-cycles with rank $r = \sigma$. Moreover the ranks r of non-bounding k-cycles satisfy the rank conditions I, II, III as is readily seen. Hence the preceding group g is a maximal group of non-bounding k-cycles, and statement (1) of the theorem is proved.

We continue with the following lemma.

LEMMA 8.1. *A necessary and sufficient condition that a linkable k-cap u be of space type is that on M*

$$u \not\sim 0 \qquad [\text{mod}_\circ\, F < a(u)]. \tag{8.8}$$

Any k-cycle in the a-class of a space k-cap is canonical and non-bounding, and conversely any canonical non-bounding k-cycle is a space cap.

Suppose u is a space k-cap. Let v be a k-cycle in the a-class of u. If (8.8) were false $v \sim 0 \text{ mod}_\circ F < a(u)$. There would then exist a k-cycle w below$_\circ$ $a(u)$ such that $v - w \sim 0$. But $v - w$ would be in the a-class of v and hence of u contrary to the hypothesis that u is a space k-cap. Hence the condition (8.8) is necessary.

To prove the condition (8.8) sufficient we again let v be a k-cycle in the a-class of u. It follows from (8.8) that

$$v \not\sim 0 \qquad [\text{mod}_\circ\, F < a(u)] \tag{8.9}$$

and a fortiori that v is non-bounding. Thus u is a space cap. The condition (8.8) is accordingly sufficient.

The second statement in the lemma follows at once from the first.

To establish (2) in Theorem 8.6 we assign ranks $b(u)$ to space k-caps u setting $b(u) = a(u)$. It is clear that the ranks $b(u)$ satisfy the first three rank conditions.

We shall prove that the group g in (8.7) is a maximal group of space k-caps.

Observe that $g(\sigma)$ in (8.7) is a group of space k-caps with cap limit t where $\sigma = (s, t) = (\infty, t)$. The group $g(\sigma)$ is a maximal group of space k-caps with cap limit t; for there is a canonical non-bounding k-cycle in the a-class of each space k-cap u. Moreover the cap limits of space k-caps satisfy the first three rank conditions so that the groups $g(\sigma)$ must sum to a maximal group of space k-caps as stated.

The proof of the theorem is complete.

The three preceding theorems are combined and extended in the following theorem.

THEOREM 8.7. *If the k-cap limits are well-ordered for all integers k, an arbi-*

trary maximal group M_k of k-caps is a-isomorphic with a maximal group of k-caps of the form

$$g_k + \overset{*}{g}_{k-1} + Q_k \tag{8.11}$$

where g_k is a suitably chosen maximal group of ambiguous k-cycles, $\overset{*}{g}_{k-1}$ is a group of k-caps $*$-isomorphic with g_{k-1}, and Q_k is an arbitrary maximal group of canonical non-bounding k-cycles. The group Q_k is isomorphic with the k^{th} homology group. When the dimensions p_k and R_k of M_k and Q_k respectively are finite the following relations hold

$$\begin{aligned} p_0 &\geqq \overset{*}{R}_0 , \\ p_1 - p_0 &\geqq R_1 - R_0 , \\ p_2 - p_1 + p_0 &\geqq R_2 - R_1 + R_0 , \\ &\cdots\cdots\cdots\cdots \end{aligned} \tag{8.12}$$

There is a group L_k of the form (8.11) which is a maximal group of k-caps by virtue of the three preceding theorems. It follows from Theorem 5.4 that L_k is a-isomorphic with M_k. The group Q_k is a maximal group of non-bounding k-cycles in accordance with Theorem 8.6 and so is isomorphic with the k^{th} homology group. If γ_k is the dimension of g,

$$p_k - R_k = \gamma_k + \gamma_{k-1} \qquad (\gamma_{-1} = 0, k = 0, 1, \cdots)$$

provided these dimensions are finite. The inequalities (8.12) follow at once.

III. Critical Points

9. **Homotopic critical points.** As stated in the introduction we distinguish between differential, homotopic, combinatorial, and essential critical points. Of these types of points the last three are topological. To define homotopic critical points we shall admit deformations of the following type.

Deformations. Let E be a subset of M. We shall consider deformations D of E which replace a point p initially on E at the time $t = 0$ by a point

$$q = q(p, t) \qquad (p \subset E; 0 \leqq t \leqq \tau)$$

on M at the time t, where t varies on the closed interval $(0, \tau)$. We shall suppose that τ is a positive constant and that $q(p, t)$ is a continuous point function of its arguments when t is on $(0, \tau)$ and p is restricted to any compact subset of E. The curve $q = q(p, t)$ obtained by holding p fast and varying t on $(0, \tau)$ will be termed the *trajectory* T defined by p. If a point q precedes a point r on a trajectory T, q will be termed an *antecedent* of r.

We shall say that a deformation D of a set S admits a *displacement* function $\delta(e)$ on S if whenever q is an antecedent of r such that $qr > e > 0$ then $F(q) - F(r) > \delta(e)$ where $\delta(e)$ is a positive single-valued function of e.

Consider the special case in which F is continuous, S compact, and D is such that $F(q) > F(r)$ whenever q is an antecedent of r distinct from r. In this case D always admits a displacement function $\delta(e)$. If however F is merely

lower semi-continuous more stringent conditions are needed to establish the existence of a displacement function, as the following example shows.

Example 9.1. Let the space M consist of the points (x, y) in the square $0 \leqq x \leqq 1$, $0 \leqq y \leqq 1$. On this square let $F = -xy$ except when $y = 0$. When $y = 0$ let $F = -x$. The function F is lower semi-continuous on M. Let E be the points of M on the y axis and let D be a deformation in which each point of E moves parallel to the x axis at a unit velocity for a unit of time. The deformation D is such that $F(q) > F(r)$ whenever q is an antecedent of r distinct from r. Nevertheless the deformation D admits no displacement function on E.

F-deformations. A deformation D of E which admits a displacement function on each compact subset of E will be termed an *F-deformation* of E, and E will be said to be F-deformable under D. We shall say that E is F-deformable *definitely onto* F_c (written onto$_\circ$ F_c) if E is F-deformable onto a point set on $F \leqq c - e$ where e is a positive constant.

In the special case where F is continuous and M compact, a point p at which $F(p) = c$ will be said to be *homotopically ordinary* if there exists a neighborhood of p on $F \leqq c$ which is F-deformable onto$_\circ$ F_c. In the general case where F is lower semi-continuous and M not necessarily compact the definitions are more complex. The preceding definition is a special case of a definition which we shall now give.

A limiting pair (p, c) (see §6) will be termed *homotopically ordinary* if there exists a neighborhood of p on $F \leqq c$ which is F-deformable onto$_\circ$ F_c. A limiting pair (p, c) which is not homotopically ordinary will be termed *homotopically critical,* and p will then be termed *semi-critical* at the level c. A point p which is semi-critical at the level $F(p)$ will be termed *critical.*

A point p may be semi-critical at different levels c. For example the point $(0, 1)$ in Example 1.1 is semi-critical at the levels $c = 0$ and $c = 1$. When F is continuous a semi-critical point p is a critical point. We shall presently apply our theory to the calculus of variations and shall thereby be concerned with a function F which is lower semi-continuous but not in general continuous. The most important special case in this variational theory is the locally convex case and in this case it will appear that each semi-critical point p is a critical point p.

Differential critical points. Let F be a function $F(x_1, \cdots, x_n)$ of class C^2 in an open region R of the space (u). We term a point $(x) = (a)$ *differentially critical* if all the first partial derivatives of F vanish at (a). Otherwise we term (a) *differentially ordinary.*

We shall show that a point which is differentially ordinary is homotopically ordinary. To that end let

$$x_i = x_i(a_1, \cdots, a_n, t) = x_i(a, t) \qquad (i = 1, 2, \cdots, n)$$

be the trajectory defined by the differential equation

$$\frac{dx_i}{dt} = -F_{x_i}$$

with the initial conditions $x_i(a, 0) = a_i$. Upon replacing the point (a) by the point $[x(a, t)]$ we obtain an F-deformation of the points (a) neighboring any ordinary point (a^0) of R, provided t be restricted to a sufficiently small interval $(0, \tau)$. With the aid of this F-deformation it is seen that each point of R which is differentially ordinary is homotopically ordinary. Hence a homotopic critical point of $F(x)$ is a differential critical point. The converse is not true as the differential critical point $x = 0$ of the function x^3 proves.

The preceding section has affirmed the existence of cycle-caps. The existence of cycle-caps implies the existence of semi-critical points in accordance with the following fundamental theorem.

THEOREM 9.1. *Corresponding to each cycle-cap u there is at least one homotopic semi-critical point p at the level $a(u)$ on each carrier of u.*

We shall prove this theorem with the aid of two lemmas. The first lemma is as follows.

LEMMA 9.1. *Let D be an F-deformation of an e-neighborhood p_e of a point p. Let A be a compact set. There exists an F-deformation D' of A which is identical with D for points initially on $p_{e/3}$.*

The deformation D' of the lemma can be defined as follows. Suppose that the given deformation D is defined for a time interval $(0, \tau)$. For the deformation D' the time t shall vary on the same interval $(0, \tau)$. Points of A initially on $p_{e/3}$ shall be deformed in D' as in D, while points of A not on $p_{2e/3}$ shall be held fast during D'. For points q of A for which

$$\frac{e}{3} < qp \leqq \frac{2e}{3} \tag{9.1}$$

we define D' as follows. Let t_q divide the time interval $(0, \tau)$ in the ratio inverse to that in which qp divides the interval $(e/3, 2e/3)$. In the deformation D', points q of A which satisfy (9.1) initially shall be deformed as in D until the time t reaches t_q and shall be held fast thereafter. It follows that D' deforms points initially on A continuously. Further A admits a displacement function under D', namely the displacement function belonging to the intersection of A with $p_{2e/3}$ under the deformation D. The deformation D' is thus an F-deformation of A, and the proof of the lemma is complete.

Our second lemma is as follows.

LEMMA 9.2. *If A is a compact set on $F \leqq a$ which contains no semi-critical points at the level a there exists an F-deformation Δ of A onto$_0$ F_a.*

If there are no points on A in limiting pairs of the form (p, a) the set A lies on$_0$ F_a and we can take Δ as the identity.

If (p, a) is a limiting pair with p on A, (p, a) is homotopically ordinary by hypothesis. There accordingly exists an F-deformation D_p of a spherical neighborhood V_p of p on $F \leqq a$ onto $F \leqq a - e_p$ where e_p is a positive constant. Let U_p and W_p be respectively spherical neighborhoods of p with radii one-third and one-sixth that of V_p. The set π of points p on A in limiting pairs (p, a) is closed and hence compact. There accordingly exists a finite set of the neigh-

borhoods W_p which cover π. Let W_{p_i}, $i = 1, \cdots, n$, be a set of such neighborhoods.

Let B_1 be an F-deformation of A which deforms the points of A initially on U_{p_1} as in D_{p_1}. Such a deformation exists by virtue of the preceding lemma. Let $G_1 = A$. Proceeding inductively with $i > 1$ let G_i be the final image of G_{i-1} under B_{i-1} and let B_i be an F-deformation of G_i which deforms the points of G_i initially on U_{p_i} as in D_{p_i}. Such a deformation exists by virtue of the preceding lemma.

We introduce the product deformation

$$\Delta = B_n \cdots B_1, \tag{9.2}$$

applying this deformation to points of A. We understand thereby that a point q of A is first deformed under B_1 into a final image $B_1 q$. The point $B_1 q$ is then deformed under B_2 into a final image $B_2 B_1 q$. The deformation Δ is the result of the application of the deformations $B_1, \cdots, B_n$ in this order. It is clear that Δ deforms points of A continuously.

We shall show that Δ is an F-deformation. To that end let $\delta_i(e)$ be the displacement function for G_i under B_i. Let q be an antecedent of r under Δ on a trajectory λ whose initial point is p. If $qr > e$ at least one of the deformations B_i must have displaced an image of p on λ between q and r, a distance greater than e/n. Hence

$$F(q) - F(r) > \min \delta_i \left(\frac{e}{n}\right) \qquad (i = 1, \cdots, n). \tag{9.3}$$

The deformation Δ thus admits a displacement function. We denote this displacement function by $\delta(e)$.

Let ρ be the minimum of the radii of the neighborhoods W_{p_i} and let β be the minimum of the numbers e_{p_i} and $\delta(\rho)$. We continue with a proof of the following statement.

(i). *The set A is F-deformed under Δ onto$_o$ F_a.*

It will be sufficient to show that the final images on W_{p_i} of points of A lie on $F \leqq a - \beta$. To that end we need only prove that the final images on W_{p_i} of points of G_i under $B_n \cdots B_i = \Delta_i$ lie on $F \leqq a - \beta$. Points of G_i will be divided into two classes.

Class I. Points on U_{p_i}.

Class II. Points not on U_{p_i}.

Points of G_i in Class I are deformed onto $F \leqq a - e_p \leqq a - \beta$ under B_i and hence under Δ_i. Points of G_i in Class II whose final images ζ under Δ_i are on W_{p_i} are displaced under Δ_i a distance at least ρ so that F is thereby decreased at least $\delta(\rho)$. These points ζ thus lie on $F \leqq a - \beta$, and the proof of statement (i) is complete.

Lemma 9.2 follows from statement (i) and the theorem follows from Lemma 9.2.

Spaces locally F-connected. We shall give a brief description of a special hypothesis of interest. The space M will be said to be *locally F-connected*

for the order n if corresponding to n, an arbitrary point p on M, and an arbitrary positive constant e, there exists a positive constant δ with the following property. For $c \geqq F(p)$ any singular n-sphere on $F \leqq c$ (the continuous image on $F \leqq c$ of an ordinary n-sphere) on the δ-neighborhood p_δ of p is the boundary of a singular $(n + 1)$-cell on $F \leqq c + e$ and on p_e. The hypothesis of local F-connectedness is not implied by the ordinary form of local connectedness at least if F is merely lower semi-continuous.

In the locally convex case of our abstract variational theory we shall be dealing with a space Ω and function F defined on Ω such that Ω is locally F-connected. In this application the values $F = 1$ will correspond to curves of infinite length.

Let c be any constant less than 1 and d be any constant greater than c. Let $R_k(c, d)$ denote the number of k-cycles in a maximal linear set of k-cycles on $F \leqq c$ not homologous to zero on $F \leqq d$. We state the following theorem.

THEOREM 9.2. *If the space M is locally F-connected for orders at most $k + 1$, the numbers $R_k(c, d)$ are finite for $c < 1$ and the inferior cycle limits of non-bounding k-cycles cluster at most at $c = 1$.*

The proof of the first statement of the theorem while not difficult involves too great detail to be presented here. The essential ideas are those in analogous proofs in Lefschetz, ref. 3. The inferior cycle limits of non-bounding k-cycles can have no cluster value $b < 1$. Otherwise corresponding to any constant c such that $b < c < 1$ there would exist an infinite sequence u^n of non-bounding k-cycles on $F \leqq c$ with distinct inferior cycle limits less than c. The ranks of the cycles u^n would be different so that the cycles u^n would generate a group of non-bounding k-cycles of infinite dimensions. It would follow that $R_k(c, 1)$ would be infinite contrary to the first statement in the theorem.

The proof of the theorem is complete.

10. **Homotopic critical sets.** We are concerned in this section with homotopic critical points. With this understood the word homotopic will ordinarily be omitted.

By a *complete critical set* ω at the level c is meant the set of all semi-critical points at the level c. It follows from the definition of a semi-critical point that ω is closed. The set ω is compact if $c < 1$ since the domain $F \leqq c$ is then compact and ω is a closed subset of this domain. Recall that the components of a point set A are the maximal connected subsets of A and are closed in A. The components of A are disjunct and sum to A. We shall term any sum of components of ω a *critical set* σ at the level c.

If σ is at a positive distance from $\omega - \sigma$, σ will be termed *separate*. A neighborhood of σ which is at a positive distance from $\omega - \sigma$ will also be termed separate.

Let R be a subset of M. If we regard R as a space M, waiving the condition that the subspace $F \leqq c$ of R be compact when $c < 1$, a k-cap on R has a new meaning dependent on R and will be specially designated as a *k-cap rel R*.

The following theorem is useful.

THEOREM 10.1. *Let σ be a critical set at the level c with a separate neighborhood U.*

(a). *If u is a k-cap with cap limit c rel U, u is a k-cap rel M.*

(b). *The k-cap u is c-homologous on U to a k-cap v on an arbitrarily small neighborhood of σ.*

We shall prove Theorem 10.1 with the aid of two lemmas, the first of which follows. In these lemmas the e- or η-neighborhood of a point set X will be denoted by X_e or X_η respectively.

LEMMA 10.1. (i). *Let U be a separate neighborhood of a critical set σ at the level c. Let b be a constant less than c, and η an arbitrary positive constant. A k-cap u with cap limit c and with carrier κ on $U + F_b$ is c-homologous on κ to a k-cap on U_η.*

(ii). *Let the complete critical set ω at the level c be the sum of critical sets $\sigma^i (i = 1, \cdots, n)$ with neighborhoods U^i at positive distances from one another. Let u^i be a k-cycle mod$_0$ $F < c$ on $F \leqq c$ and U^i, such that the sum $u^1 + \cdots + u^n = u$ is c-homologous to 0 on $F_b + \Sigma U^i$. Then u^i is c-homologous to 0 on U^i_η for each i.*

We shall first prove (i).

Let u_p be the p^{th} component of u, and H_p the relative algebraic homology connecting u_p and u_{p+1} in the definition of u. Of the cells on the complement of U we drop each k-cell of u_p and each $(k + 1)$-cell in H_p. Let u'_p and H'_p be the k-chain and relative homology which thereby replace u_p and H_p respectively. The sequence u'_p together with the homologies H'_p defines a k-cap u', c-homologous to u on κ. If m is a sufficiently large integer the subsequence u'_p for which $p \geqq m$ together with the corresponding homologies H'_p will define a k-cap z on U_η. Moreover z and u' and hence z and u will be c-homologous on κ.

The proof of (i) is complete.

We come to the proof of (ii).

We set $\Sigma U^i = V$ and let e be an arbitrary positive constant less than one-half the minimum distance between pairs of neighborhoods U^i. By hypothesis v is c-homologous to 0 on $V + F_b$. From the algebraic homologies which define this c-homology we drod all $(k + 1)$-cells which are on the complement of V and all algebraic homologies whose norms are greater than $e/2$. The resulting algebraic homologies will hold on $V_{e/2}$ and we infer that u is c-homologous to 0 on V_e. It then follows from the choice of e that the cycle u^i is c-homologous to 0 on the corresponding neighborhood U^i_e.

The proof of (ii) is complete.

We come to a deformation lemma.

LEMMA 10.2. *Let κ be a compact subset of $F \leqq c$ on a separate neighborhood of a critical set σ at the level c. Corresponding to an arbitrary positive constant e there exists a positive constant $b < c$ and an F-deformation T of κ onto $\sigma_e + F_b$ in which points of κ which have been displaced a distance at least e lie on F_b.*

Set $e = 3\rho$. Let κ_1 and κ_2 be respectively the subsets of κ not on σ_ρ and $\sigma_{2\rho}$.

There are no semi-critical points on κ_1 at the level c. It follows from Lemma 9.2 that there exists an F-deformation Δ of κ_1 onto $F < c - \eta$ where $\eta > 0$. The methods used in the proof of Lemma 9.1 make it clear that the deformation Δ properly modified will yield an F-deformation T of κ in which $\kappa - \kappa_1$ is held fast while κ_2 is deformed as in Δ. Let $\delta(e)$ be the displacement function for κ under T.

Points of κ are on κ_2 or $\sigma_{2\rho}$. Points of κ on κ_2 are F-deformed under T onto $F < c - \eta$. Points of κ on $\sigma_{2\rho}$ are F-deformed on $\sigma_{3\rho} = \sigma_e$ or else displaced a distance at least ρ, and hence deformed onto $F \leqq c - \delta(\rho)$. Under T points of κ which have been displaced a distance at least e will lie on $F \leqq c - \delta(e)$. If b is less than c and greater than $c - \eta$, $c - \delta(\rho)$ and $c - \delta(e)$ the lemma holds as stated.

We shall now prove Theorem 10.1 beginning with a proof of (a).

Let θ be a carrier of u on U and $F \leqq c$. We assume (a) false. It follows that u is c-homologous to 0 with homology carrier $\kappa \supset \theta$ on $F \leqq c$. Corresponding to the complete critical set ω, and to the compact set κ Lemma 10.2 affirms the existence of a constant $b < c$ and an F-deformation T of κ onto $\omega_e + F_b$. Under T, θ is deformed on $\theta_e + F_b$. Hence u is c-homologous to 0 on $\omega_e + \theta_e + F_b$.

We distinguish between the cases $\omega = \sigma$ and $\omega \neq \sigma$. If $\omega \neq \sigma$ we apply Lemma 10.1 (ii) with $n = 2$, setting

$$\sigma^1 = \sigma, \quad \sigma^2 = \omega - \sigma, \quad u^1 = u, \quad u^2 = 0, \quad U^1 = \sigma_e^1 + \theta_e, \quad U^2 = \sigma_e^2,$$

taking e so small that $\overline{U}^1 \subset U$, and U^2 is at a positive distance from U^1. We see that u is c-homologous to 0 on $F_b + \Sigma U^i$ and infer that u is c-homologous to 0 on U. If $\omega = \sigma$ we again apply Lemma 10.1 (ii) with $n = 1$. In both cases we conclude that u is c-homologous to 0 on U.

From this contradiction we infer the truth of (a).

We turn to a proof of (b).

Let e be so small a positive constant that σ_{2e} lies on U. The k-cap u lies on U and it follows from Lemma 10.2 that u is c-homologous to a k-cap w on $\sigma_e + F_b$. Since σ_e is a separate neighborhood of σ it follows from Lemma 10.1 (i) that w is c-homologous to a k-cap v on σ_{2e} (setting $\eta = e$). Thus $u - v$ is c-homologous to 0. But from our choice of e, $u - v$ is on U. It follows from (a) that $u - v$ is c-homologous to 0 on U. But v is on σ_{2e} and e may be arbitrarily small.

Statement (b) follows directly.

A k-cap u with cap limit c will be said to be *associated* with a critical set σ if there is a k-cap in the a-class of u with a carrier on an arbitrary neighborhood of σ on $F \leqq c$. With this understood we state the following corollary.

COROLLARY 10.1 *If σ is a critical set at the level c with a separate neighborhood U, a maximal group g of k-caps rel U with cap limits c is a maximal group of k-caps associated with σ.*

It follows from (a) in Theorem 10.1 that each k-cap u of g is a k-cap rel M

and from (b) that u is "associated" with σ. Thus g is a group of k-caps associated with σ. That g is maximal among k-caps v associated with σ is seen as follows. The k-cap v is in the a-class of a k-cap u rel U by virtue of its association with σ. But g is maximal among k-caps rel U with cap limit c, so that u and hence v is in the a-class of a k-cap of g. The corollary thus holds as stated.

It follows from Theorem 5.1 that any two maximal groups g of k-caps with cap limit c are a-isomorphic. The following theorem adds an important mode of determination of a particular maximal group of k-caps.

THEOREM 10.2. *Let the complete critical set ω at the level c be the sum of critical sets σ^i $(i = 1, \cdots, n)$ with neighborhoods V^i at positive distances from one another and let g^i be a maximal group of k-caps rel V^i with cap limit c. The groups g^i admit a direct sum g and this sum is a maximal group of k-caps with cap limit c.*

We begin by proving the following statement.

(α). *If v^i is an element of g^i the sum $v = v^1 + \cdots + v^n$ is c-homologous to 0 only if for each i, v^i is c-homologous to 0 on V^i.*

If v is c-homologous to 0 it follows from (a) Theorem 10.1 that v is c-homologous to 0 on ΣV^i. Since the neighborhoods V^i are at positive distances from one another we infer that v^i is c-homologous to 0 on V^i, and the proof of (α) is complete.

It remains to prove statement (β).

(β). *A k-cap v with cap limit c is c-homologous to a sum $v^1 + \cdots + v^n$ of elements v^i in the respective groups g^i.*

It follows from Theorem 10.1 (b) that v is c-homologous to a k-cap on an arbitrarily small neighborhood of ω and hence to a sum of cycles u^i mod$_\circ$ $F < c$ on $F \leqq c$ and on the respective neighborhoods V^i. But each such cycle u^i is c-homologous on V^i to an element v^i of g^i in accordance with the definition of g^i, and (β) follows immediately.

Theorem 10.2 follows from (α) and (β).

The dimension of a maximal group of k-caps associated with a critical set σ will be termed the k^{th} *type number* of σ. It follows from Corollary 10.1 that the k^{th} type number of a separate critical set σ is the dimension of a maximal group of k-caps with cap limit c relative to a separate neighborhood of σ. We thus have the following corollary of Theorem 10.2.

COROLLARY 10.2. *The k^{th} type number of the complete critical set ω is the sum of the k^{th} type numbers of the critical sets $\sigma^1, \cdots, \sigma^n$.*

A k-cap u with cap limit c which is associated with a critical set σ will be said to *belong* to σ if u is associated with no proper critical subset of σ. The basic theorem here is as follows.

THEOREM 10.3. *A k-cap u with a cap limit $c < 1$ belongs to at least one critical set at the level c.*

We shall not use this theorem and accordingly omit its proof.

Non-degenerate differential critical points. We shall consider the important case in which M is compact and in which there is a neighborhood of each point

p of M homeomorphic with a region U of a euclidean n-space of rectangular coördinates (x). Neighboring any point q of U we admit any system of coördinates obtainable from the coördinates (x) by a non-singular transformation $z_i = z_i(x)$ with $z_i(x)$ of class C^2, and we admit only such coördinate systems neighboring q. A space M of this character will be said to be *locally of class* C^2. We suppose that our function F equals a function $\varphi(x)$ of class C^2 corresponding to each admissible set of coördinates (x).

A differential critical point (x_0) of $\varphi(x)$ will be termed a *differential critical point* of F. The point (x_0) will be termed degenerate if the Hessian of φ at (x_0) is zero. We *assume* here that the differential critical points of F are non-degenerate. As in the introduction the index of a differential critical point (x_0) shall be the *index* of the quadratic form whose coefficients make up the Hessian of φ at (x_0). From the hypothesis of non-degeneracy it follows that the differential critical points are isolated, and hence finite in number.

As in §9 the trajectories orthogonal to the manifolds φ constant in the local coördinate spaces (x) will serve to prove that homotopic critical points are differential critical points. A proof of the following theorem is more difficult.

THEOREM 10.4. *A non-degenerate differential critical point σ is a homotopic critical set and the j^{th} type number of σ equals the Kronecker delta δ^j_k where k is the index of σ.*

We suppose that σ is represented by the point $(x) = (0)$ in the coördinate system (x) and that in this system $F \equiv \varphi(x)$. We also suppose that $\varphi(0) = 0$, waiving the condition that $0 \leqq F \leqq 1$. Setting

$$f(x) = -x_1^2 - \cdots - x_k^2 + x_{k+1}^2 + \cdots + x_n^2,$$

we introduce a preliminary transformation as follows:

(a). *If r is a sufficiently small positive constant the space A defined by the conditions $\varphi \leqq 0$ and $x_i x_i < r^2$, and the space B defined by the conditions $f \leqq 0$ and $x_i x_i < r^2$ admit a homeomorphism in which the subspaces $f < 0$ and $\varphi < 0$ correspond.*

It follows from methods developed in detail on pp. 169–170 of M1 that the space $\varphi \leqq 0$, $x_i x_i = r^2$ and the space $f \leqq 0$, $x_i x_i = r^2$ admit a homeomorphism in which the subspaces $f < 0$ and $\varphi < 0$ correspond. A use of the radial trajectories of pp. 156–157 of M1 then gives an easy proof of (a).

We shall make use of Corollary 10.1, determining the j^{th} type number of σ, if such a number exists, by determining the dimension μ of a maximal group of j-caps with cap limit 0 relative to a separate neighborhood of σ on $\varphi \leqq 0$. The space A is such a neighborhood if r is sufficiently small, since each homotopic critical point is a differential critical point and the only differential critical point on A at the level 0 is σ if r is sufficiently small. It follows from (a) that μ equals the dimension of a maximal group of j-caps with cap limit 0 rel B with f replacing φ. We continue, replacing A by B and φ by f.

If $k = 0$, B consists of the point σ, and the theorem is immediate. We therefore suppose that $k > 0$. When $k > 0$ we shall prove the following:

(b). *The homology groups of the subspace* $f < 0$ *of* B *are isomorphic with those of the* $(k - 1)$*-sphere.*

(c). *The only non-null connectivity of* B *is* $R_0 = 1$.

(d). *Each* j*-cap* u *rel* B *with* $a(u) = 0$, *is non-linkable.*

(e). *The only* j*-caps* u *rel* B *with* $a(u) = 0$ *are* k*-dimensional, and a maximal group of such* k*-caps has the dimension* 1.

The space B and its subspace $f < 0$ can be respectively deformed on themselves onto the disc

$$x_1^2 + \cdots + x_k^2 \leqq \frac{r^2}{2}, \qquad x_{k+1} = \cdots = x_n = 0 \tag{10.1}$$

and its spherical boundary S, holding the disc and S fast respectively. The truth of (b) and (c) follows.

Proof of (d). If there were a linkable j-cap e rel B with $a(u) = 0$ there would be a j-cycle v in the a-class of u. But any j-cycle on B is homologous on B to a cycle on S, and $f = -r^2/2$ on S. Hence v cannot be a j-cap with cap limit 0 and u cannot be linkable.

Proof of (e). If u is a j-cap rel B with $a(u) = 0$, u is non-linkable. By definition then, βu is on $f < 0$ and is there non-bounding. It follows from (*b*) that $j = 1$ or k. If $k = 1$, the first statement in (*e*) is true. If $k > 1$, I say that $j = k$. For if j were 1 it would follow from the connectedness of the subspace $f < 0$ of B when $k > 1$ that $\beta u \sim 0$ $\text{on}_\circ$ this subspace, contrary to the fact that u is non-linkable. Hence $j = k$ in all cases and the proof of the first statement in (e) is complete.

To prove the second statement in (e) recall that the subspace $f < 0$ of B contains a $(k - 1)$-cycle u which is non-bounding $\text{on}_\circ$ this subspace but bounding on B. Hence there is at least one k-cap u rel B with $a(u) = 0$.

A maximal set of k-caps rel B with $a(u) = 0$ will contain at most one k-cap. To prove this let u and v be two such k-caps. It follows from (*b*) that there exists a proper homology of the form

$$\delta_1 \beta u + \delta_2 \beta v \sim 0 \qquad (\text{on}_\circ f < 0). \tag{10.2}$$

The relative k-cycle $\delta_1 u + \delta_2 v$ cannot be a k-cap at the level 0 since it would then be linkable in accordance with (10.2). Hence $\delta_1 u + \delta_2 v$ is not a k-cap at the level 0, and the proof of (e) is complete.

We can conclude that σ is a homotopic critical point and set at the level 0. For no other point on A at the level 0 is a homotopic critical point as noted previously. But 0 is a cap limit rel A as the preceding paragraphs show. It follows from Theorem 9.1 that σ is a homotopic critical point. Since σ is an isolated critical point it must also be a critical set, and the proof of the theorem is complete.

A compact space M locally of class C^2 has finite connectivities (cf. Lefschetz, *Deformations*, l.c.), and if M is locally n-dimensional the connectivities of order greater than n are null. If F has at most non-degenerate critical points

the number of critical points is finite and it follows from Theorems 10.4, 10.2 and 5.3 that a maximal group of k-caps will have a dimension m_k equal to the number of critical points of index k. In particular $m_j = 0$ if $j > n$. Referring to the relations (8.12) we have the following theorem.

THEOREM 10.5. *Let F be a function of class C^2 on an n-dimensional compact space M locally of class C^2. If the differential critical points of F are non-degenerate the numbers p_k of critical points with index k and the connectivities R_i of M satisfy the relations* (0.2) *of the introduction.*

Essential critical points. A homotopic critical set σ to which at least one k-cap belongs will be termed *homotopically essential* and in such a case a semi-critical point of σ will be termed *homotopically essential.* The class of homotopic semi-critical points B thus has a subclass C of points which are termed essential. Thus $B \supset C$. That the class C may be a proper subclass of B is shown by the following example.

Example 10.2. Let M be a curve $y = f(x)$ in the (x, y)-plane defined as follows:

$$f(x) = \sin x \qquad (-\pi \leqq x \leqq 0),$$

$$f(x) = 0 \qquad (0 \leqq x \leqq \pi),$$

$$f(x) = \sin x \qquad (\pi \leqq x \leqq 2\pi).$$

On M we set $F = (y + 1)/2$ so that $0 \leqq F \leqq 1$. The points of the curve at which $0 \leqq x \leqq \pi$ are homotopically critical and form a critical set but not an essential critical set.

We have already seen that the class A of differential critical points in problems where F is differentiable is such that $A \supset B$. We thus have the relation

$$A \supset B \supset C, \tag{10.3}$$

but B may be a proper subset of A and C a proper subset of B.

IV. ABSTRACT VARIATIONAL THEORY

11. **Length.** Let N be a space of points $p, q, r, \cdots$ with a metric which in general is not symmetric. That is to each ordered pair p, q of points of N there shall correspond a number denoted by pq and called the distance from p to q. These distances shall satisfy the conditions:

I. $pp = 0$,

II. $pq > 0$ if $p \neq q$,

III. $pq \leqq pr + rq$.

The relation III is not to be written $pq \leqq pr + qr$ inasmuch as this would imply that $pq = qp$ upon setting $r = p$. The spaces N satisfying I–III include those with a symmetric distance. For various applications, for instance to the

calculus of variations, it is not desirable to restrict ourselves to the symmetric case. The non-symmetric case is illustrated by the integral

$$J = \int (x\dot{y} + \sqrt{\dot{x} + \dot{y}})\, dt \qquad (0 < x < 1). \tag{11.1}$$

The value of J along a curve in general is changed when the sense of integration along the curve is reversed. Curves which are extremals in one sense need not be extremals in the other sense. The absolute minimum of J from one point to another is a distance which satisfies I–III but which is not in general symmetric.

Let $| pq |$ denote the maximum of pq and qp. We shall call $| pq |$ the *absolute distance* between p and q. With $| pq |$ as a distance N becomes a metric space with a symmetric distance function. We shall denote this new space by $| N |$. We shall use the metric of N when defining length. For all other purposes, in particular in defining point functions on N or $| N |$, or general topological properties, we shall use the metric of $| N |$. One reason for this is that pq does necessarily approach zero when qp approaches zero, unless indeed we add further conditions to the conditions I–III.

We understand that a function $f(p_1, \cdots, p_n)$ of points $(p_1, \cdots, p_n)$ on $| N |$ is continuous at $(q_1, \cdots, q_n)$ if $f(p_1, \cdots, p_n)$ tends to $f(q_1, \cdots, q_n)$ as a limit as max $| p_i q_i |$ tends to zero. This is the familiar concept of continuity on $| N |$. A function is called continuous on a point set A if continuous at each point of A. We note the following:

(a). *The function pq is a uniformly continuous function of p and q on $| N |$.*

Postulate III implies that for points p, q, p', q',

$$pq \leqq pp' + p'q' + q'q. \tag{11.2}$$

If we assume that $| pp' | < \delta$ and $| qq' | < \delta$ then (11.2) implies that $pq < p'q' + 2\delta$. Upon interchanging p and p', q and q' we find that pq and $p'q'$ differ by less than 2δ so that (a) is true.

For each value of t on a closed interval $(0, a)$ let $q(t)$ be a point on $| N |$. The function $q(t)$ is termed *continuous* at t if the absolute distance $| q(t') \, q(t) |$ tends to zero as t' tends to t. If $q(t)$ is continuous at each point t on $(0, a)$ the set of points $q(t)$ taken in the order of the corresponding values of t will be termed a parameterized curve λ (written p-curve λ). We say that λ is the continuous image on $| N |$ of the line segment $(0, a)$. In general curves on $| N |$ will be denoted by Greek letters α, β, η, λ, etc., while points on $| N |$ will be denoted by letters p, q, r, etc.

The length of λ. Let t_i be a set of values on $(0, a)$ such that $0 = t_0 < t_1 < \cdots < t_n = a$, and let $p_0, p_1, \cdots, p_n$ be the corresponding points on λ. The sets t_i and p_i will be termed *partitions* of λ of *norm* δ where δ is the maximum of the differences $t_{i+1} - t_i$. We term

$$\sum = \sum_{i=0}^{n-1} p_i p_{i+1},$$

a sum *approximating the length* of λ and take the length L of λ as the least upper bound of the sums Σ for all partitions of λ. The proof of the following theorem can be readily supplied by the reader familiar with the ordinary theory of length in euclidean spaces. Explicit use of course must be made of postulate III.

THEOREM 11.1. *The length of λ is the limit of any sequence Σ_n of approximating sums Σ corresponding respectively to partitions of λ whose norms tend to zero as n becomes infinite.*

Fréchet distance. Let η and ζ be two p-curves given by the respective equations

$$p = p(t) \qquad (0 \leqq t \leqq a), \tag{11.3}$$

$$q = q(u) \qquad (0 \leqq u \leqq b). \tag{11.4}$$

Let H be a sense preserving homeomorphism between the closed intervals $(0, a)$ and $(0, b)$. The homeomorphism between η and ζ defined by H will be termed *admissible*. Set

$$d(H) = \max \mid p(t)q(u) \mid,$$

where t and u correspond under H. Then the Fréchet distance $\eta\zeta$ between η and ζ is the greatest lower bound of the numbers $d(H)$ as H ranges over all admissible homeomorphisms between η and ζ.

We observe that $\eta\zeta = \zeta\eta \geqq 0$. If moreover η, ζ, λ are three p-curves,

$$\eta\lambda \leqq \eta\zeta + \zeta\lambda. \tag{11.5}$$

To establish (11.5) let H be an admissible homeomorphism between η and ζ under which a point p of η corresponds to a point q of ζ. Let K be an admissible homeomorphism between ζ and λ in which q of ζ corresponds to r of λ. Let KH be the resultant transformation of η into λ carrying p of η into r of λ. From the relation $\mid pr \mid \leqq \mid pq \mid + \mid qr \mid$ we find that

$$d(KH) \leqq d(H) + d(K). \tag{11.6}$$

Relation (11.5) follows from (11.6).

By the *space* of the *p-curves* will be meant a space in which the elements are p-curves and the distance between elements the Fréchet distance.

We continue with a proof of a well-known theorem.

THEOREM 11.2. *The length of a p-curve η is a lower semi-continuous function of η in the space of p-curves η.*

We are to show that if $a < L(\eta)$ there exists a positive δ such that $L(\zeta) > a$ whenever $\eta\zeta < \delta$.

Let b be a constant between a and $L(\eta)$. By virtue of the definition of L there exists a partition $p_0, p_1, \cdots, p_n$ of η such that

$$\sum p_i p_{i+1} > b \qquad (i = 0, 1, \cdots, n-1). \tag{11.7}$$

We choose δ so that $2n\delta < b - a$. If $\eta\zeta < \delta$ there exists a homeomorphism T between η and ζ such that corresponding points have an absolute distance less than δ. In particular if q_i is the correspondent of p_i under T the points q_i determine a partition of ζ such that $| p_i q_i | < \delta$ and hence

$$q_i q_{i+1} > p_i p_{i+1} - 2\delta \qquad (i = 0, 1, \cdots, n-1). \tag{11.8}$$

With the aid of (11.8) we see that

$$L(\zeta) \geqq \sum q_i q_{i+1} > \sum p_i p_{i+1} - 2n\delta.$$

It follows from (11.7) and our choice of δ that

$$L(\zeta) \geqq \sum q_i q_{i+1} > b - (b - a) = a,$$

and the proof is complete.

If $p_1, \cdots, p_n$ is an arbitrary sequence of points on $| N |$ it follows from postulate III that

$$p_1 p_n \leqq p_1 p_2 + p_2 p_3 + \cdots + p_{n-1} p_n .$$

This inequality together with the definition of length leads to the following theorem.

THEOREM 11.3. *The length of a p-curve joining two points p and q on* $| N |$ *is at least pq.*

μ-Length. Let η and ζ be the p-curves given by (11.3) and (11.4) respectively. We shall say that ζ is *derivable from* η if there exists a continuous non-decreasing function $u = u(t)$ which maps the closed interval $(0, a)$ onto the closed interval $(0, b)$ and under which $p(t) \equiv q[u(t)]$. Two p-curves which are derivable from the same p-curve η will be said to be *p-equivalent*. The two p-curves η and ζ will be said to be *identical*, $\eta \equiv \zeta$, if in the representations (11.3) and (11.4) $a = b$ and $p(t) \equiv q(t)$. To show that p-equivalence is transitive it will be sufficient to verify the truth of the following statement.

(a). *Corresponding to each arbitrary p-curve* η *there exists a p-curve* φ_η *from which* η *is derivable such that* $\varphi_\alpha \equiv \varphi_\beta$ *whenever the p-curves* α *and* β *are p-equivalent.*

If a p-curve of the form (11.3) has finite length, the length $s(t)$ of η from the point $t = 0$ to a general point t is a continuous non-decreasing function of t. If $t = t(s)$ is the function inverse to $s(t)$, $p[t(s)]$ is a single-valued continuous point function of s for s on the interval $0 \leqq s \leqq s(a)$. We set $q(s) \equiv p[t(s)]$. The curve $q = p(t)$ is derivable from the curve $q = q(s)$ upon setting $s = s(t)$. We term $q = q(s)$ the s-curve *determined* by η. If ζ is derivable from η it is clear that the s-curve determined by ζ is identical with the s-curve determined by η. For curves with finite length the curve φ_η corresponding to η in statement (a) could be taken as the s-curve determined by η. It follows that p-equivalence is transitive for p-curves of finite length.

But for curves with infinite length there are no s-curves. Nevertheless it is possible to replace length s by what will be termed μ-length and to associate with each p-curve η a p-curve φ_η such that statement (a) is satisfied without

exception. The parameter of these new curves will be denoted by μ and the curves themselves will be termed μ-curves.[13] These μ-curves will not only serve as the p-curves φ_η of (a) but will have other important properties as follows.

(b) *The parameter μ on a p-curve φ_η varies from 0 to a value $\mu(\eta)$. The function $\mu(\eta)$ is continuous in η in the space of p-curves η and never exceeds the diameter of η.*

(c). *The condition that η and ζ be p-equivalent is equivalent to the condition that the μ-curves φ_η and φ_ζ be identical and to the condition $\eta\zeta = 0$.*

(d). *Let α be a parameter which ranges from 0 to 1 inclusive. Let the point on φ_λ at which $\mu = \alpha\mu(\lambda)$ be denoted by $q(\lambda, \alpha)$. Then $q(\lambda, \alpha)$ is continuous in its arguments.* See ref. 16. Page 441.

The following is a particular consequence of (d).

(e). *Corresponding to a preassigned (fixed) p-curve η and an arbitrary positive e there exists a positive δ such that if $\eta\delta < \delta$, points on φ_η and φ_ζ which divide φ_η and φ_ζ in the same ratio with respect to variation of μ have an absolute distance at most e.*

12. **The space** Ω. A class of p-equivalent p-curves will be termed a *curve*. The p-curves of a given class λ have a 0-distance from each other. The distance $\lambda\tau$ between two curves λ, τ shall be defined as the distance $\eta\zeta$ between any two p-curves η and ζ in the respective classes λ and τ. It is clear that the distance $\lambda\tau$ is independent of the curves η and ζ chosen to represent λ and τ respectively. It follows from property (c) of μ-curves in §11 that $\lambda\tau = 0$ if and only if the classes λ and τ are identical. Curves λ, τ, ω, $\cdots$ accordingly form a symmetric metric space M of the same character as the space M of §1.

Let a and b be two points on $|N|$. The set of curves joining a to b will form a subspace $\Omega(a, b)$ of the space M of all curves on $|N|$. We shall investigate the properties of $\Omega(a, b)$.

Recall that a subset A of a metric space is termed complete if every Cauchy sequence in A converges to a point in A, and conditionally compact if every infinite sequence of A contains a Cauchy subsequence. (See Hausdorff, l.c.) If A is conditionally compact and complete A is compact.

Apart from trivial cases where $a = b$, or the set $\Omega(a, b)$ is empty, $\Omega(a, b)$ is not conditionally compact. In fact if $\Omega(a, b)$ contains a curve λ it is possible to define an infinite sequence of curves λ_n each of which joins a to b and is defined by a point $p(t)$ which varies continuously on λ but is such that the sequence λ_n contains no Cauchy subsequence.

We have denoted the length of a p-curve by L. We note that the value of L will be the same for each p-curve of a class λ of p-curves which are p-equivalent.

---

[13] Morse, M., *A special parameterization of curves*, Bulletin of the American Mathematical Society, 42 (1936). This paper contains a proof of statements (a), (b), (c), (d) and a definition of μ-curves. This definition is a modification of a definition of a function of sets given by H. Whitney and applied by Whitney to families of simple non-intersecting curves. See Whitney, H., *Regular families of curves*, Annals of Mathematics, 34 (1933), pp. 244–270.

This common value of L will be termed the length of λ and again denoted by L. We understand that L may be infinite.

We shall prove the following theorem. See Fréchet ref. 17, page 441.

THEOREM 12.1. *If* $|N|$ *is compact and* c *is a finite constant, the subspace* $L \leqq c$ *of* $\Omega(a, b)$ *is compact.*

We shall begin with a proof of statement (i).

(i). *If* $|N|$ *is complete,* $\Omega(a, b)$ *is complete.*

Let λ_1, λ_2, be a Cauchy sequence Λ of curves of $\Omega(a, b)$. Given $e > 0$, there exists an integer N so large that $\lambda_i\lambda_j < e$ whenever $i > N$ and $j > N$. We choose a subsequence of Λ as follows. Let η_1 be chosen from Λ so that the distance of η_1 from every curve λ_n following η_1 in Λ is less than $\frac{1}{2}$. Proceeding inductively, for $i > 1$ choose η_i from curves λ_j following η_{i-1} in Λ so that if λ_m follows η_i in Λ,

$$\eta_i\lambda_m < 2^{-i}.$$

The sequence $\eta_1, \eta_2, \cdots$ is a subsequence of Λ with the property that

$$\eta_i\,\eta_{i+1} < 2^{-i}. \tag{12.1}$$

Let $\tau_1, \tau_2, \cdots$ be a sequence of homeomorphisms of which τ_i is a homeomorphism between η_i and η_{i+1} in which corresponding points p_i and p_{i+1} have an absolute distance less than 2^{-i}. Such homeomorphisms exist by virtue of (12.1). For $j > i$ the homeomorphisms $\tau_i\,\tau_{i+1}\cdots\tau_{j-1}$ taken in the order written yield a homeomorphism τ_{ij} between η_i and η_j by virtue of which corresponding points p_i and p_j have an absolute distance

$$|p_ip_j| < \frac{1}{2^i} + \frac{1}{2^{i+1}} + \cdots + \frac{1}{2^{j-1}} < \frac{1}{2^{i-1}}. \tag{12.2}$$

Suppose η_1 is represented by the p-curve $p = p_1(t)$, $0 \leqq t \leqq 1$. By virtue of the homeomorphism τ_{1j} a p-curve $p = p_j(t)$ representing η_j is uniquely determined, in such manner that $p_i(t)$ and $p_j(t)$ correspond under τ_{ij}. Relation (12.2) now takes the form

$$|p_i(t)\,p_j(t)| < \frac{1}{2^{i-1}}. \tag{12.3}$$

It follows from the completeness of $|N|$ that the sequence of points $p_i(t)$ converges uniformly with respect to t to a continuous point limit $p(t)$. The sequence η_i thus converges to the curve represented by $p = p(t)$, and the proof of (i) is complete.

We continue with a proof of statement (ii).

(ii). *If* $|N|$ *is conditionally compact, the subspace* $L \leqq c$ *of* $\Omega(a, b)$ *is conditionally compact.*

We shall make use of the n-fold product $|N|^n$ of $|N|$. The distance between two points

$$(p) = (p_1, \cdots, p_n), \qquad (q) = (q_1, \cdots, q_n)$$

of $|N|^n$ shall be defined as the maximum of the distances $|p_i q_i|$ and shall be denoted by $(p)(q)$.

Let Λ^0 be an infinite sequence of curves $\lambda_1^0, \lambda_2^0, \cdots$ on $L \leqq c$. Let $e_1, e_2, \cdots$ be a sequence of positive constants tending to zero. Let $n_1 > 0$ be an integer so large that the division of the curve λ_i^0 into parts of equal variation of length by the ordered set of points

$$(p^i) = (p_1^i, \cdots, p_{n_1}^i) \qquad (i = 1, 2, \cdots) \tag{12.4}$$

yields segments of λ_i^0 with absolute diameters less that $e_1/3$. An integer n_1 exists since we are concerned with curves of length at most c.

Since $|N|$ is conditionally compact, $|N|^{n_1}$ is conditionally compact. The points (p^i) in the space $|N|^{n_1}$ will thus have a Cauchy subsequence (r^j), $j = 1, 2, \cdots$, and we can assume that the sequence (r^j) has been so chosen that

$$(r^i)(r^k) < \frac{e_1}{3} \qquad (i, k = 1, 2, \cdots). \tag{12.5}$$

Let λ_i^1 be the curve of Λ^0 of which (r^i) is a partition. It follows from (12.5) and from the fact that the set of points (r^i) divide λ_i^1 into segments of absolute diameter less than $e_1/3$ that

$$\lambda_j^1 \lambda_k^1 < e_1 \qquad (j, k = 1, 2, \cdots). \tag{12.6}$$

The sequence λ_i^1 has been selected from the sequence λ_i^0. We repeat this process replacing e_1 by e_2 and λ_i^0 by λ_i^1, thereby obtaining a subsequence λ_i^2 of λ_i^1. More generally we obtain a subsequence $\lambda_1^n, \lambda_2^n, \cdots$ of the sequence $\lambda_1^{n-1}, \lambda_2^{n-1}, \cdots$ of such a nature that

$$\lambda_i^n \lambda_k^n < e_n \qquad (i, k = 1, 2, \cdots).$$

The sequence

$$\lambda_1^1, \lambda_2^2, \lambda_3^3, \cdots \tag{12.7}$$

is a subsequence of Λ^0 and has the property that

$$\lambda_n^n \lambda_m^m < e_r \qquad (\text{if } n, m > r). \tag{12.8}$$

It is accordingly a Cauchy subsequence.

The proof of (ii) is complete.

We can now prove the theorem. It follows from (i) that the space $\Omega(a, b)$ is complete. The subspace $L \leqq c$ of $\Omega(a, b)$ is closed since L is lower semi-continuous. Hence this subspace is complete. The theorem follows from (ii).

Since F is lower semi-continuous Theorem 12.1 has the following corollary.

COROLLARY. *If $\Omega(a, b)$ is not empty there exists at least one curve λ of $\Omega(a, b)$ whose length (possibly infinite) is a minimum*[14] *among lengths of curves of $\Omega(a, b)$.*

[14] See Tonelli, *Fondamenti di calcolo delle Variazioni* I and II. References will here be found to the work of Hilbert on the absolute minimum.

If H is a component (in the point set sense) of $\Omega(a, b)$ there is at least one curve ζ of H whose length (possibly infinite) is a minimum among lengths of curves of H.

A class of curves of $\Omega(a, b)$ continuously deformable on $| N |$ among curves of $\Omega(a, b)$ into a given curve of $\Omega(a, b)$ will be termed a *homotopy* class. Among curves joining two fixed points a and b on a torus for example there are infinitely many homotopy classes, and on a torus these homotopy classes are identical with the components of $\Omega(a, b)$. Under the condition of local convexity to be developed in the next chapter we shall see that the homotopy classes are identical with the components of Ω. Our corollary will then affirm the existence of at least one curve in each homotopy class K of Ω with length a minimum among lengths of curves of K.

Let $H^k(a, b)$ be the k^{th} homology group of $\Omega(a, b)$. Let $x \ominus y$ be read "x is simply isomorphic with y." We see that $H^k(a, b) \ominus H^k(b, a)$. We shall extend this result by proving the following theorem.

THEOREM 12.2. *If $| N |$ is arcwise connected,*

$$H^k(a, b) \ominus H^k(c, d), \tag{12.9}$$

where (a, b) and (c, d) are arbitrary pairs of points on $| N |$.

We shall begin by proving statement (i).

(i). *If there is an arc on $| N |$ joining b to c then $H^k(a, b) \ominus H^k(a, c)$.*

Let η be a curve joining b to c on $| N |$. Let λ be a curve joining a to b. By the *extension* of λ by η we mean the curve obtained by tracing λ and η successively. We denote this extension by $(\lambda)_\eta$ and observe that $(\lambda)_\eta$ joins a to c.

Let a_k be an oriented k-cell on $\Omega(a, b)$. By the extension $(a_k)_\eta$ of a_k by η we mean the oriented k-cell on $\Omega(a, c)$ obtained by replacing each vertex of a_k by its extension by η, understanding of course that a vertex of a_k is a curve of $\Omega(a, b)$. Let $z_k = \delta^i a_k^i$ be an algebraic k-chain on $\Omega(a, b)$. By the *extension* $(z_k)_\eta$ of z_k is meant the k-chain

$$(z_k)_\eta = \delta^i (a_k^i)_\eta \qquad [\text{on } \Omega(a, c)]. \tag{12.10}$$

We observe that

$$\beta(z_k)_\eta = (\beta z_k)_\eta, \tag{12.11}$$

that is boundary and extension operators are commutative. If u_j is the j^{th} component of a Vietoris cycle u, the algebraic j-cycles $(z_j)_\eta$ form a new Vietoris k-cycle $(u)_\eta$ on $\Omega(a, c)$. This follows in part from (12.11) and in part from the fact that the distance between two curves of $\Omega(a, b)$ equals the distance between their extensions on $\Omega(a, c)$, so that norms are not altered by the process of extension.

Let η' be the curve obtained by tracing η in the opposite sense. If z_k is an algebraic k-cycle of norm e on $\Omega(a, b)$, set

$$z_k^* = [(z_k)_\eta]_{\eta'}. \tag{12.12}$$

We see that z_k^* is an algebraic k-cycle of norm e on $\Omega(a, b)$.

We shall define a deformation D which continuously deforms z_k into z_k^* on $\Omega(a, b)$.

To that end let η be given in the form $p = p(\tau)$ with $0 \leqq \tau \leqq 1$ and let η_t denote the subarc of η on which $0 \leqq \tau \leqq t \leqq 1$. Under D suppose that an initial point λ (i.e. curve) of $\Omega(a, b)$ is replaced at the time t by the point (i.e. curve)

$$[(\lambda)_{\eta_t}]_{\eta_t'} \qquad (0 \leqq t \leqq 1)$$

of $\Omega(a, b)$ where η_t' represents η_t taken in the opposite sense. The deformation D carries z_k into z_k^* as stated. Hence

$$z_k^* \sim_{e+\delta} z_k \qquad [\text{on } \Omega(a, b)], \tag{12.13}$$

where e is the norm of z_k^* and z_k and δ is an arbitrary positive constant.

We return to the proof of (i) *and set up an isomorphism between the groups $H^k(a, b)$ and $H^k(a, c)$.*

To a k-cycle z of $\Omega(a, b)$ corresponds the cycle $(z)_\eta$ of $\Omega(a, c)$. It follows from (12.11) that k-cycles which are homologous on $\Omega(a, b)$ thereby correspond to k-cycles which are homologous on $\Omega(a, c)$. Each homology class of $H^k(a, b)$ is thus mapped onto a homology class of $H^k(a, c)$. We designate this mapping by T. That this mapping is a homeomorphism follows at once from the definition of the operation of extension by η. It remains to prove that T is an isomorphism.

The mapping T leads to each class of $H^k(a, c)$. If u is a k-cycle on $\Omega(a, c)$, $[(u)_{\eta'}]_\eta$ is a k-cycle on $\Omega(a, c)$, homologous to u as we see from (12.13) upon interchanging (a, b) and (a, c). But $(u)_{\eta'}$ is on $\Omega(a, b)$. Thus T leads to each class of $H^k(a, c)$.

To show that the mapping T is one-to-one we have merely to show that the null class in $H^k(a, c)$ corresponds to the null class in $H^k(a, b)$. Suppose a k-cycle on $H^k(a, c)$ is the extension $(z)_\eta$ of a k-cycle z on $H^k(a, b)$. If $(z)_\eta \sim 0$, we have

$$0 \sim (z_n)_{\eta'} \sim z \qquad [\text{on } \Omega(a, b)].$$

Hence T is an isomorphism as stated, and the proof of (i) is complete.

To prove the theorem we note that the operations replacing b by c in $H^k(a, b)$, or of interchanging a and b lead to groups isomorphic with $H^k(a, b)$. We can thus successively replace (a, b) by (a, c), (c, a), (c, d) and obtain a group $H^k(c, d)$ isomorphic with $H^k(a, b)$.

The proof of the theorem is complete.

We state the following corollary of the theorem.

Corollary. *The connectivities of $\Omega(a, b)$ are topological invariants of $| N |$.*

Whether or not the connectivities of $\Omega(a, b)$ are determined by arithmetic topological invariants of $| N |$ now known is an interesting unsolved problem. It seems probable that they are not so determined.

If $| N |$ is a space homeomorphic with an n-sphere with $n > 1$, the connectivities R_k of $\Omega(a, b)$ are 1 if $k \equiv 0 \bmod n - 1$, otherwise null. For example, for $n = 2$ and 3 respectively, the numbers R_k form the sequences $1\ 1\ 1 \cdots$ and $1\ 0\ 1\ 0 \cdots$. See M1, Theorem 15.1, p. 247.

The function F. Let η be an arbitrary curve of $\Omega(a, b)$ and let $L(\eta)$ be the length of η. We set

$$F(\eta) = \frac{L(\eta)}{1 + L(\eta)} \tag{12.14}$$

understanding thereby that $F = 1$ when L is infinite. The function F is lower semi-continuous with $0 \leqq F \leqq 1$. For $c < 1$ the subdomains $F \leqq c$ of $\Omega(a, b)$ are compact in accordance with Theorem 12.1. A curve λ of $\Omega(a, b)$ will be called a *homotopic pseudo-extremal* or *extremal* respectively if λ is a homotopic semi-critical point or a critical point of F respectively.

Our theory of functions F defined on spaces M is now applicable.

We turn to the development of conditions under which a pseudo-extremal is always an extremal, that is conditions under which a critical pair (p, c) is of the form $[p, F(p)]$.

V. Locally Convex Spaces

13. **Elementary arcs.** In this chapter we shall suppose that the space of §11 is conditioned by two new postulates IV and V. In stating V we shall need the notion of an *elementary arc.* If p and q are two distinct points of $| N |$, an arc λ joining p to q will be termed *elementary* if it has the following two properties. It is simple, that is it is defined by a one-to-one continuous image $p(t)$ of a line segment $0 \leqq t \leqq a$; a point r shall lie on λ if and only if

$$pq = pr + rq. \tag{13.1}$$

Postulates IV and V are as follows.

IV. *The space $| N |$ shall be compact and connected.*

V. *There shall exist a positive constant ρ such that whenever $0 < pq \leqq \rho$ there exists at least one elementary arc which joins p to q.*

The elementary arc postulated in V will be denoted by (pq). We shall term V the postulate of *local convexity* and ρ the *norm of convexity.* We shall restrict ourselves to elementary arcs (pq) for which $pq \leqq \rho$.

The postulates I–V are independent as can be readily shown. They are satisfied on any ordinary closed manifold without singularity and with a sufficiently regular local representation. The elementary arcs are then geodesics, and geodesic arcs issuing from a point p and cut off at a sufficiently small distance ρ from p cover a neighborhood of p in a one-to-one continuous manner, p alone excepted. On $| N |$ however the elementary arcs issuing from a point p cover a neighborhood of p in general multiply as examples would show.

We regard two elementary arcs as identical if they are similarly sensed and consist of the same points. With this understood we have the following theorem.

Theorem 13.1. *There is at most one elementary arc joining a point p to a point q.*

Let λ and ζ be elementary arcs joining p to q. Let r be a point on λ between p and q. Then $pq = pr + rq$. By virtue of the definition of an elementary

arc this implies that r lies on ζ as well as λ, and the proof of Theorem 13.1 is complete.

The following theorem is of importance.

THEOREM 13.2. *Any subarc of an elementary arc* (pq) *is an elementary arc.*

We shall begin by proving statements (1) and (2) below. In these statements and elsewhere we shall make the special convention that the elementary arc joining a point p to p is the point p itself.

(1). *If r is a point on (pq), then the arc pr on (pq) is elementary.*

(2). *If r is a point on (pq), then the arc rq on (pq) is elementary.*

To establish (1) we first note that $pr \leqq \rho$ since $pr + rq = pq \leqq \rho$. It follows from V that there is an elementary arc (pr) joining p to r. Let s be any point on (pr). Then $pr = ps + sr$. Upon adding rq to both sides of this equality we find that

$$pr + rq = ps + sr + rq. \tag{13.2}$$

Since r is on (pq) we can replace the left member of (13.2) by pq. In the right member it follows from postulate III that $sr + rq \geqq sq$ so that (13.2) takes the form

$$pq \geqq ps + sq. \tag{13.3}$$

Upon comparing (13.3) with postulate III we infer that $pq = ps + sq$ so that s must lie on (pq).

We have yet to show that s is on the arc pr of (pq). To this end let (pq) and (pr) be respectively the one-to-one continuous images of the segments $0 \leqq t \leqq 1$ and $0 \leqq x \leqq 1$. Given any value of x on $0 \leqq x \leqq 1$ there is determined a unique point P on (pr) and, since P lies on (pq) as we have just seen, a unique value of t. We denote this value of t by $t(x)$. It is readily seen that the relation $t = t(x)$ is one-to-one and continuous for x on $0 \leqq x \leqq 1$. Hence (pr) must coincide with the arc pr of (pq), and (1) is proved.

The proof of (2) is similar to that of (1) interchanging the rôles of p and q in the proof of (1) and appropriately changing the order of the points involved. To establish the theorem let p, r, s, q be points on (pq) in the order written. By virtue of (1) the arc ps of (pq) is elementary. It follows from (2) that the arc rs on (ps) is elementary, and the proof is complete.

Theorem 13.2 leads to the following.

THEOREM 13.3. If $r = r(\tau)$, $0 \leqq \tau \leqq 1$, *defines an elementary arc joining p to q, the distance $pr(\tau)$ is a continuous increasing function of τ.*

The continuity of $pr(\tau)$ is a consequence of the continuity of $r(\tau)$ and the continuity of pr as a function of p and r. To show that $pr(\tau)$ is an increasing function of τ we note that

$$pr(\tau') = pr(\tau) + r(\tau)r(\tau') \qquad (0 \leqq \tau < \tau' \leqq 1), \tag{13.4}$$

so that $pr(\tau') > pr(\tau)$.

We are able to prove a theorem which is much stronger than Theorem 13.3.

To state this theorem let (pq) be an elementary arc with variable end points p, q. Let t be a number between 0 and pq inclusive. Let the point r on (pq) at the distance t from p be denoted by $f(p, q, t)$. Our theorem is the following.

THEOREM 13.4. *The point function $f(p, q, t)$ is continuous for $pq \leqq \rho$ and $0 \leqq t \leqq pq$.*

Let (p_n, q_n, t_n) be a sequence of admissible sets (p, q, t) which tend to a set (p^0, q^0, t^0) as a limit set. Set $r_n = f(p_n, q_n, t_n)$. Since $|N|$ is compact there is a subsequence of the points r_n with a limit point r^0. For simplicity we assume that the sequence r_n itself converges to r^0. We shall prove that

$$r^0 = f(p^0, q^0, t^0). \tag{13.5}$$

Upon letting n become infinite in the relation $p_nq_n = p_nr_n + r_nq_n$ we infer that

$$p^0q^0 = p^0r^0 + r^0q^0. \tag{13.6}$$

We also have the relations

$$p^0r^0 = \lim_{r=\infty} p_nr_n = \lim_{n=\infty} t_n = t^0. \tag{13.7}$$

From (13.6) and (13.7) we see that r^0 is the point on the elementary arc p^0r^0 which is at the distance t^0 from p^0. Hence (13.5) holds, and the theorem is true.

The following theorem depends upon the compactness of $|N|$ and postulates I–III.

THEOREM 13.5. *The absolute distance $|pq|$ approaches zero uniformly with pq, and the absolute diameter of any curve approaches zero uniformly with its length.*

Suppose the first affirmation of the theorem false. There will then exist a sequence of pairs of points (p_n, q_n) whose absolute distances are bounded from zero but whose distances p_nq_n tend to zero as n becomes infinite. There will exist at least one limit pair p^*, q^* of the pairs p_n, q_n. For such a pair $p^*q^* = 0$ so that $p^* = q^*$. But $|p^*q^*| \neq 0$ since the distances $|p_nq_n|$ are bounded from zero. From this contradiction we infer that $|pq|$ approaches zero uniformly with pq.

The concluding affirmation of the theorem follows from the definition of length and the first statement in the theorem.

Recall that a metric space M is termed *separable* if there is a subset of points of M whose closure is M and which is at most countably infinite. We state the following theorem.

THEOREM 13.6. *Under postulates* I–V *the space $\Omega(a, b)$ is separable.*

The space $|N|$ is compact and therefore separable. Let (q) be a set of points of $|N|$ at most countably infinite with closure $|N|$. We suppose that the points a and b belong to (q). By a broken extremal ζ is meant a curve composed of a finite succession of elementary arcs. The end points of these elementary arcs will be called the *vertices* of ζ. Let H be the set of all broken extremals which join a to b and whose vertices belong to (q). It is clear that

the set H is at most countably infinite. Corresponding to an arbitrary curve λ of $\Omega(a, b)$ there is a broken extremal ζ within a prescribed distance e of λ. We suppose ζ so chosen that the lengths of its elementary arcs are at most $\rho/2$. Let the successive vertices of ζ be designated by $a, p_1, \cdots, p_n, b$. We replace the vertices p_i by points q_i of (q) so near the respective points p_i that the successive points $a, q_1, \cdots, q_n, b$ can still be joined by elementary arcs, and so near p_i that the distance of the resulting broken extremal η from ζ is at most e. The curve η belongs to the set H, and $\lambda\eta \leqq 2e$.

The proof of the theorem is complete.

14. **Metric extremals.** We begin this section with the following theorem.

THEOREM 14.1. *The length of an elementary arc (pq) equals pq and is a proper minimum relative to the lengths of all curves which join p to q.*

Let $p_1, \cdots, p_n$ be a set of points which appear in the order written on (pq). Any segment of an elementary arc is an elementary arc. It then follows inductively for $n = 1, 2, \cdots$ that

$$pq = pp_1 + p_2p_3 + \cdots + p_nq.$$

We infer that the length of (pq) equals pq.

Let λ be any curve joining p to q. It follows from Theorem 11.3 that $L(\lambda) \geqq pq$. It remains to show that $L(\lambda) > pq$ when λ is not the elementary arc (pq).

The proof falls into two cases.

Case I. The curve λ contains a point s not on (pq).

Case II. Each point of λ is on (pq), but $\lambda \not\equiv (pq)$.

In Case I we see that $L(\lambda) \geqq ps + sq > pq$, and the proof is complete.

In Case II there must be distinct points r and s on λ and on (pq) which appear in the order rs on (pq) but in the order sr on λ. Then

$$L(\lambda) \geqq ps + sr + rq > ps + rq > ps + sq = pq,$$

and the proof of the theorem is complete.

A *metric extremal* λ shall be the continuous image on $|N|$ of a closed or open interval T of the t-axis such that every inner point t of T is an inner point of a closed subinterval of T whose image on λ is an elementary arc. A curve which is not a metric extremal will be termed *metrically ordinary*. We continue with the following theorem.

THEOREM 14.2. *Let a and b be two points on $|N|$. Under the hypothesis of local convexity a curve λ of minimum length joining a to b is a metric extremal.*

The successive points of a sufficiently fine partition of λ can be joined by elementary arcs thereby forming a curve of finite length. Hence the length of λ must be finite. Moreover each arc η of λ with a length less than ρ must be an elementary arc. For if p and q are the end points of η, $pq \leqq L(\eta) < \rho$ in accordance with Theorem 14.1 so that p and q can be joined by an elementary arc (pq). If η is not an elementary arc $pq < L(\eta)$ by virtue of Theorem 14.1

and λ can be shortened by replacing η by (pq). It follows that λ is a metric extremal, and the proof is complete.

Homotopic extremals and pseudo-extremals were defined in §12 before the hypothesis of local convexity was introduced, while metric extremals were defined after this hypothesis was introduced. The principal theorem of this section is the following. It includes Theorem 14.2 as a special case.

THEOREM 14.3. *Under the postulates* I–V *the class of homotopic extremals on* $\Omega(a, b)$ *is identical with the class of homotopic pseudo-extremals on* $\Omega(a, b)$ *and each such extremal is a metric extremal.*

The proof of this theorem will require the introduction of a lemma and a basic deformation θ. We proceed with the necessary definitions and proofs.

Admissible sets π. Let λ be a curve on $| N |$ with end points a and b. If r is a sufficiently large integer, there is a set of $r - 1$ points on λ dividing λ into a sequence of arcs

$$\pi = (\lambda_1, \cdots, \lambda_r) \tag{14.1}$$

on each of which the distance from the initial point to the terminal point is at most ρ. A set π of curves λ_i obtained in this manner from a curve λ joining a to b on M will be termed *admissible*. We term λ_i the i^{th} component of π. Corresponding to an admissible set π we denote the curve λ which π determines by $\lambda(\pi)$.

Let π be an admissible set (λ_i). We denote the elementary arc which joins the initial point of λ_i to the terminal point of λ_i by λ_i^*. The set of curves λ_i^* will then be denoted by π^*. The set π^* is determined by π. The curve $\lambda(\pi^*)$ is a broken extremal.

The following lemma is a broad generalization of Osgood's Theorem.[15] In it we refer to the function $F(\lambda)$ defined in §13.

LEMMA 14.1. *For admissible sets* π *which possess a fixed number* r *of components and for which the distance of* $\lambda(\pi)$ *from* $\lambda(\pi^*)$ *is not less than a positive constant* e *the value of* F *on* $\lambda(\pi)$ *exceeds the value of* F *on* $\lambda(\pi^*)$ *by an amount which is bounded away from zero.*

In proving this lemma admissible sets π will function as points in a metric space. The distance between two admissible sets π,

$$\pi = (\lambda_1, \cdots, \lambda_r), \qquad \pi' = (\lambda_1', \cdots, \lambda_r')$$

with the same number of components shall be defined as the maximum distance $\lambda_i\lambda_i'$ and denoted by $\pi\pi'$.

---

[15] See Tonelli, l.c.

[16] Cf. Fréchet Sur une représentation paramétrique intrinsèque de la courbe continue la plus générale Journal de Mathématique t. IV, 1925, pp. 281–297. This very interesting representation could not be used in the present paper because it lacked the property (d) of our representations.

[17] Fréchet *Sur quelques points du calcul fonctionnel*, Rendiconti del Circolo Matematico di Palermo, vol. 22 (1906), pp. 1–74.

We consider a fixed positive integer r and observe that $L[\lambda(\pi^*)] \leqq r\rho$ for all admissible sets π with r components. Hence there is a constant m dependent on r such that

$$F[\lambda(\pi^*)] \leqq m < 1.$$

Let k be an arbitrary number between m and 1. We divide admissible sets π with r components into two classes I and II according as

$$\text{I. } F[\lambda(\pi)] > k, \qquad \text{II. } F[\lambda(\pi)] \leqq k.$$

We then introduce the function

$$H(\pi) = F[\lambda(\pi)] - F[\lambda(\pi^*)].$$

If π is in Class I,

$$H(\pi) > k - m > 0. \tag{14.2}$$

We turn to Class II. There is a constant c dependent on k such that each component λ_i of an admissible set π in Class II has a length at most c. The set of all curves on $| N |$ with lengths at most c forms a compact space G (the proof of this is similar to the proof of Theorem 12.1). Let G^r denote the product space in which G is an r-fold factor. The space G^r is compact. For r fixed, the sets π which are admissible and belong to Class II, and for which the distance $\lambda(\pi)\lambda(\pi^*) \geqq e$ form a closed subset G_0 of G^r.

The function $H(\pi)$ is lower semi-continuous. For $F(\lambda)$ is lower semi-continuous in λ, $\lambda(\pi)$ varies continuously with admissible sets π and $F[\lambda(\pi^*)]$ is a continuous function of admissible sets π. By virtue of Theorem 14.1, $H(\pi)$ is positive on G_0. At some point π_0 of G_0, $H(\pi)$ then takes on a positive absolute minimum. That is

$$H(\pi) \geqq H(\pi_0) > 0 \qquad (\pi \subset G_0).$$

Combining this result with (14.2) we find that

$$H(\pi) \geqq \min [k - m, H(\pi_0)] > 0$$

for the sets π of the lemma.

The proof of the lemma is complete.

We shall define an F-deformation $\theta(\eta, m)$ of the neighborhood of an arbitrary curve η of $\Omega(a, b)$. To that end we shall need a representation of curves λ of $\Omega(a, b)$ in terms of μ-length. See §11. Let $\mu(\lambda)$ be the μ-length of λ and let α be a parameter which ranges from 0 to 1 inclusive. Let the point q on λ at which the μ-length of λ equals $\alpha\lambda(\mu)$ be designated by

$$q = q(\lambda, \alpha). \tag{14.3}$$

The function $q(\lambda, \alpha)$ is conditioned by properties (b), (d) and (e) of μ-length as given in §11.

The deformation $\theta(\eta, m)$. Let m be a positive integer so large that η admits a partition into m successive subarcs with the following properties. The diam-

eter of each of these arcs shall be less that $\rho/2$. If η is not an extremal at least one of these subarcs shall not be an elementary arc. Let V be a neighborhood of η on $\Omega(a, b)$ and λ a curve on V. The curve λ shall be divided into m successive subarcs on which the variation of the parameter α in $q(\lambda, \alpha)$ in (14.3) shall equal its variation on the corresponding subarcs of η. It follows from property (e) of μ-length in §11 that the neighborhood V can be chosen so small that the diameter of each of the m-subarcs of λ is less than $\rho/2$. We suppose V so chosen.

Let ω be the i^{th} subarc of λ. We shall define θ for ω, understanding that the deformation θ of λ is obtained by simultaneously applying θ to each of the subarcs ω of λ. In the deformation θ the time t shall range from 0 to 1 inclusive. Let $t_n = 2^{-n}$, $n = 0, 1, 2, \cdots$. We shall first define θ for the times t_n.

For $t = t_n$ the deformation θ shall replace ω by a broken extremal ω_n whose end points are the end points of ω and whose successive vertices divide ω into 2^n successive segments of equal variation of the parameter α of λ. Let $(\alpha)_n$ denote the ordered set of parameter values α at the end points and points of division of ω.

Let the curve which replaces ω at the time t be denoted by ω^t. We shall now define ω^t when t is on the interval

$$t_n < t < t_{n-1}.$$

Let α' and α'' denote any pair of successive values of α in the set $(\alpha)_{n-1}$. In the set $(\alpha)_n$ there is just one value α^* of α between α' and α''. Let g denote the elementary arc of ω_n which joins the point α^* on ω to the point α'' on λ. Let t divide the interval (t_n, t_{n-1}) in an arbitrary ratio with respect to length. Let $p(t)$ denote the point on g which divides g in the same ratio with respect to length. Corresponding to each pair α', α'' chosen as above we replace the arc ξ of ω_n with end points at the points α' and α'' on ω and intermediate vertex at α^* on ω by the broken extremal with end points of ξ and one intermediate vertex $p(t)$, thereby obtaining the curve ω^t.

The definition of $\theta(\eta, m)$ is complete.

In showing that θ is an F-deformation we shall use the concept of an *adjoin* of a point set A of $|N|$. By the adjoin of A is meant the set of points on all elementary arcs whose end points lie on A. By the 2-*fold adjoin* of A is meant the adjoin of the adjoin of A. We shall make use of the following lemma.

LEMMA 14.2. *Corresponding to an arbitrary positive constant e there exists a positive constant δ such that the diameter of the 2-fold adjoin of A is at most e whenever the diameter of A is at most δ.*

This lemma is a consequence of our theorems on elementary arcs.

To prove that θ is an F-deformation of the neighborhood V of η we must establish the following:

(i). *The curve λ_t replacing λ at the time t varies continuously with λ and t for λ on any compact subset S of V and for $0 \leqq t \leqq 1$.*

(ii). *The deformation θ admits a displacement function on S.*

Let ζ denote an arbitrary point of S and τ an arbitrary value of t on the interval $0 \leqq t \leqq 1$. Regarding λ_t as defined for the moment only for curves λ on S we shall prove that λ_t is continuous in λ and t at the pair (ζ, τ). We shall distinguish two cases: I. $\tau = 0$; II. $\tau > 0$.

We first consider Case I. The point q on λ has been given the representation $q = q(\lambda, \alpha)$. The function $q(\lambda, \alpha)$ is continuous in its arguments in accordance with (d) of §11. It follows that for λ on S the 2^n successive arcs k into which ω was divided in defining θ will be uniformly small in diameter provided $n \geqq N$, and N is a sufficiently large integer. Let e be an arbitrary positive constant. It follows from Lemma 14.2 that if N is sufficiently large the 2-fold adjoin of two successive arcs k will have a diameter at most $e/2$. In defining θ, λ_t was obtained from λ by replacing (certain) pairs k', k'' of adjacent arcs k of λ by two successive elementary arcs of the 2-fold adjoin of $k' + k''$. For $t < 2^{-N}$ each of the above arcs k involved in the construction of λ_t belongs to at least a 2^N-fold partition of an arc ω. Set $\delta = 2^{-N}$. From the choice of N we infer that

$$\lambda_t\lambda \leqq \frac{e}{2} \qquad (\lambda \subset S, t < \delta). \tag{14.4}$$

We next observe that

$$\lambda_t\zeta \leqq \lambda_t\lambda + \lambda\zeta \tag{14.5}$$

so that we see from (14.4) and (14.5) that

$$\lambda_t\zeta \leqq e \qquad \text{when } [\lambda\zeta < e/2; \lambda \subset S; t < \delta].$$

The continuity of λ_t is established at (ζ, τ) when $\tau = 0$.

We turn to Case II. For $t \geqq \tau$ and sufficiently near τ, or for $t \leqq \tau$ and sufficiently near τ, the curve λ_t is a broken extremal whose vertices vary continuously with λ and t for λ on S in accordance with (d) of §11. Hence λ_t is continuous at (λ, τ) in Case II.

The proof of (i) is complete.

We shall continue with a proof of (ii).

Let λ be a curve of S. Let ζ' be an antecedent of ζ'' on the trajectory T of λ under θ. Let e be a positive constant. Setting $F(\zeta') - F(\zeta'') = \Delta F$ we must show that whenever $\zeta'\zeta'' > e$,

$$\Delta F > \delta(e) \tag{14.6}$$

where $\delta(e)$ is a positive constant which depends only on S and e.

Let ζ_n denote the image of λ at the time t_n. We shall say that the pair ζ', ζ'' is in Class 1 if for no n is ζ_n on T between ζ' and ζ'', and if $\zeta'\zeta'' > e$. We shall first establish (14.6) for pairs ζ', ζ'' in Class 1.

As constructed in the definition of θ, ζ'' is a succession of r elementary arcs ζ_i''. The curve ζ' is a succession of r curves ζ_i' whose end points are joined by the respective elementary arcs ζ_i''. We can apply Lemma 14.1 to ζ' and ζ'' identi-

fying ζ' with $\lambda(\pi)$ and ζ'' with $\lambda(\pi^*)$ and infer the existence of a function $\omega(r, e)$ such that

$$\Delta F > \omega(r, e) > 0 \qquad (\zeta'\zeta'' > e). \tag{14.7}$$

Relation (14.6) will follow from (14.7) provided the integer r in (14.7) is bounded for pairs ζ', ζ'' in Class 1. The integer r will fail to be bounded only if there are pairs ζ', ζ'' in Class 1 arbitrarily near λ, as one sees upon representing the curves λ of S in the form $q = q(\lambda, \alpha)$ of (14.3), and recalling that S is compact. But pairs ζ', ζ'' arbitrarily near λ would be arbitrarily near each other contrary to the hypothesis that $\zeta'\zeta'' > e$. We conclude that r in (14.7) is at most an integer R. We can accordingly take $\delta(e)$ in (14.6) as the minimum, say $\delta_1(e)$, of the numbers

$$\omega(1, e), \omega(2, e), \cdots, \omega(R, e).$$

Thus (14.6) holds for pairs ζ', ζ'' in Class 1.

We turn to pairs ζ', ζ'' not in Class 1.

As previously we admit pairs ζ', ζ'' such that $\zeta'\zeta'' > e$ and such that ζ' is an antecedent of ζ'' on a trajectory T emanating from a curve λ of S. If ζ' and ζ'' are not in Class 1 there exist integers p and s such that the images

$$\zeta_{p+1}, \quad \zeta', \quad \zeta_p, \quad \zeta_s, \quad \zeta'', \quad \zeta_{s-1} \tag{14.8}$$

of λ appear in the order written on T. We admit the possibility that consecutive curves in (14.8) may be identical. The function $F(\zeta)$ decreases monotonically as ζ moves along T in the sense of increasing t. Hence the left member of(14.6) is at least as great as each of the three numbers

$$F(\zeta') - F(\zeta_p), \qquad F(\zeta_p) - F(\zeta_s), \qquad F(\zeta_s) - F(\zeta'').$$

Admissible pairs ζ', ζ'' not in Class 1 will be assigned to Classes 2, 3, or 4 according as

$$2.\ \ \zeta'\zeta_p \geqq \frac{e}{3}, \qquad 3.\ \ \zeta_p\zeta_s \geqq \frac{e}{3}, \qquad 4.\ \ \zeta_s\zeta'' \geqq \frac{e}{3}.$$

Since $\zeta'\zeta'' > e$ each pair ζ', ζ'' will belong to at least one of these classes. Proceeding as under Class 1 we can infer the existence of a function $\delta_i(e)$, $i = 2, 3, 4$, such that $\Delta F > \delta_i(e)$ when ζ', ζ'' is in Class i.

For a fixed e let $\delta(e)$ denote the minimum of the numbers $\delta_i(e)$ for $i = 1, 2, 3, 4$. For this $\delta(e)$, (14.6) holds as stated, and the proof of (ii) is complete.

The deformation $\theta(\eta, m)$ is accordingly an F-deformation of the neighborhood V of η.

Proof of Theorem 14.3. Let η be a curve of $\Omega(a, b)$ and let $\theta(\eta, m)$ be a corresponding F-deformation of curves λ on a neighborhood of η. Let η_1 and λ_1 be respectively the final images of η and λ under θ. We shall begin by proving the following statement.

(i). *Corresponding to an arbitrary positive constant* e *there exists a positive constant* δ *such that for* $\lambda\eta < \delta$,

$$F(\lambda_1) \leqq F(\eta_1) + e. \tag{14.9}$$

The final image λ_1 of λ is a succession of m elementary arcs. Let $(p) = (p_0, \cdots, p_m)$ be the successive end points of these elementary arcs. Designate the set (p) determined by η_1 by (q). Let the distance $(p)(q)$ be defined as in §12 as the maximum of the distances $| p_i q_i |$. According to the construction of θ the points p_i divide λ in the same ratios with respect to μ-length as the points q_i divide η. For η fixed it follows from property (d) of μ-length in §11 that the distance $(p)(q)$ tends to zero with $\lambda\eta$. But $F(\lambda_1)$ may be regarded as a function $\varphi(p)$ of the set (p) determining (λ_1). The function $\varphi(p)$ is continuous in (p) at least if $(p)(q)$ is so small that (p) determines a sequence of elementary arcs. Hence

$$\varphi(q) \leqq \varphi(p) + e$$

provided $(p)(q)$ is sufficiently small. Statement (i) follows directly.

We continue with a proof of (ii).

(ii). *A homotopic pseudo-extremal is a homotopic extremal.*

Let η be a semi-critical point in the space $\Omega(a, b)$ at the level c. We shall show that $F(\eta) = c$. It follows from the lower semi-continuity of F that either $F(\eta) = c$ or $F(\eta) < c - e_1$ where $e_1 > 0$. If η_1 is the final image of η under θ and $F(\eta) < c - e_1$ it would follow with the aid of (14.9) that

$$F(\lambda_1) \leqq F(\eta_1) + e \leqq F(\eta) + e < c - e_1 + e \qquad (\lambda\eta < \delta). \tag{14.10}$$

If δ in (i) is sufficiently small e can be taken less than e_1 and (14.10) implies that

$$F(\lambda_1) < c - (e_1 + e) < c \qquad (\lambda\eta < \delta).$$

We can thus F-deform the neighborhood $\lambda\eta < \delta$ of η onto a domain definitely below c. The limiting pair (η, c) must then be homotopically ordinary contrary to hypothesis, and we conclude that (ii) is true.

We conclude the proof of Theorem 14.3 by establishing (iii).

(iii). *A homotopic extremal is a metric extremal.*

To prove (iii) it will be sufficient to show that if η is metrically ordinary η is homotopically ordinary. When η is metrically ordinary the m arcs into which η is initially divided in defining $\theta(\eta, m)$ are so chosen that at least one arc is not elementary. It follows that $F(\eta_1) < F(\eta)$ and upon making use of the arbitrariness of e it follows from (14.9) that θ F-deforms a sufficiently small neighborhood of η onto a domain definitely below $F(\eta)$. Thus η is homotopically ordinary.

The proof of Theorem 14.3 is complete.

15. **Properties of $\Omega(a, b)$ and of metric extremals.** In §9 we have defined locally F-connected spaces. We shall here prove the following theorem.

THEOREM 15.1. *Under postulates* I–V *the space* $\Omega(a, b)$ *is locally F-connected for all orders n.*

Let η be a curve of $\Omega(a, b)$. To establish the theorem it will be sufficient to establish the following.

(a). *Corresponding to* η *and a positive constant e there exists a positive constant* δ *and a deformation of* η_δ *on* η_e *into a curve* η_1 *on* η_e, *in which any compact subset S of* η_δ *on* $F \leqq c \geqq F(\eta)$ *is continuously deformed on* $F \leqq c + e$.

We shall apply the deformation $\theta(\eta, m)$ of the preceding section to η. Recall that the definition of θ began with the division of η into m successive arcs. If these arcs are sufficiently small in diameter and V is a sufficiently small neighborhood of η, θ will deform V on the $e/2$ neighborhood of η. We suppose θ and V so chosen. Let λ be a curve on V and let η_1 and λ_1 be final images of η and λ respectively under θ.

Each broken extremal λ_1 which is sufficiently near η_1 can be deformed into η_1 as follows. The vertex q_i of λ_1 will be joined to the corresponding vertex p_i of η_1 by an elementary arc E_i, and as t varies from 0 to 1, q_i shall be replaced by a point p_t which divides E_i in the same ratio with respect to length as that in which t divides the interval (0, 1). The broken extremal λ_1 will thereby be deformed into η_1. Denote this deformation by Λ.

Let δ be a positive constant such that $\eta_\delta < V$. Let H be the final image of η_δ under θ. The diameter of H tends to 0 with δ. We suppose δ so small that Λ deforms H on η_e and on $F \leqq c + e$.

The deformation θ applied to η_δ, followed by Λ applied to H, yields a deformation of η_δ which satisfies (*a*). Theorem 15.1 follows directly.

Referring to Theorem 9.2 we infer the following corollary of Theorem 15.1.

COROLLARY 15.1. *Under postulates* I–V *the lengths of extremals of* $\Omega(a, b)$ *which correspond to inferior cycle limits of non-bounding k-cycles, have no finite cluster value.*

Analytic functionals. We shall give a brief indication of how the preceding theory can be applied to functionals which are analytically defined. We suppose that the underlying space M is *compact, connected, locally analytic,* and *n-dimensional.* To be locally analytic M shall satisfy the conditions of a space locally of class C^2 as defined as §9, together with the condition that the transformations of coördinates $z_i = z_i(x)$ shall be defined by functions $z_i(x)$ which are analytic. In each local coördinate system we suppose that we have given an invariant function $f(x, r)$ defining an integral

$$J = \int f(x, \dot{x})\, dt$$

as in M1, p. 111. We suppose that $f(x, r)$ is analytic, positive, positive homogeneous of order 1 in the variables (r), and positive regular. See M1, p. 121. We presuppose the analyticity of M and $f(x, r)$ for the sake of simplicity, although it will be clear to the reader that less stringent conditions would suffice.

As is well known there will exist a positive constant ρ with the following

properties. The extremals of J which issue from a point p of M and on which $0 \leqq J \leqq \rho$ will form a field covering a neighborhood of p in a one-to-one manner, p alone excepted, with no conjugate points of p on the extremal arcs $0 < J \leqq \rho$. We introduce a metric N in which the distance pq equals the absolute minimum of J along rectifiable curves which join p to q. With N and ρ so defined, we proceed as in the abstract variational theory noting that conditions I–V are satisfied.

As previously we are concerned with the space $\Omega(a, b)$. The integral J defines "length" and leads as at the end of §12 to the functional F on $\Omega(a, b)$. A metric extremal will here be locally regular and analytic, and will be termed a *differential extremal*. We shall prove the following theorem.

THEOREM 15.2. *Let η be a differential extremal joining a to b on which there are k conjugate points of a, but on which b is not conjugate to a. The extremal g is a homotopic critical set whose j^{th} type number is δ_k^j.*

In proving this theorem we shall make use of a subset of curves of Ω termed *canonical curves* and defined as follows. Let

$$a, a_1, \cdots, a_{m-1}, b \tag{15.1}$$

be a sequence of points on η which divide η into m successive elementary arcs with diameters less than $\rho/2$. Let M^q be a regular analytic $(n - 1)$-manifold cutting η at the point a_q without being tangent to η at a_q. Broken extremals of Ω whose intermediate vertices lie on the successive manifolds M^q will be termed *canonical*, and a set of all such curves with vertices on neighborhoods of the respective points a_q will be denoted by S. The value of F on curves λ of S will be an analytic function $\varphi(u)$ of the local coördinates of M neighboring the respective vertices a_q. As shown in M1 (cf. p. 228) the point $(u) = (u_0)$ which determines η is a non-degenerate critical point of φ of index k. Upon referring to Theorem 10.4 we infer the truth of the following.

(α). *If S consists of canonical curves determined by points (u) sufficiently near (u_0) the differential extremal η is an isolated homotopic extremal rel S, and the j^{th} type number of η rel S equals δ_k^j.*

The curve η is isolated among differential extremals on Ω at the level $c = F(\eta)$. There are accordingly no homotopic extremals on Ω at the level c in a sufficiently small neighborhood of η, η excepted as we shall see.

We continue with a proof of statement (β).

(β). *Corresponding to the set S of (α) there exists a separate neighborhood U of η on $\Omega(a, b)$ which can be deformed into a subset of curves of S without increasing F beyond its initial value on each curve of U.*

We shall make use of the deformation $\theta(\eta, m)$ of the preceding section understanding that the m arcs into which η is initially divided in defining $\theta(\eta, m)$ are the m arcs whose vertices are given by (15.1). We take U so small that U is on the domain upon which $\theta(\eta, m)$ operates as an F-deformation. Let λ be a curve of U and μ the final image of λ under $\theta(\eta, m)$. The curve μ will be a broken extremal with intermediate vertices neighboring the respective points

a_q of η. If U is sufficiently small there will be a unique curve in S whose successive vertices lie on μ. We denote this curve by $\zeta(\mu)$.

The deformation Z. Under Z we shall deform μ into $\zeta(\mu)$. We thereby replace each intermediate vertex p of μ by a variable vertex p_t. We cause p_t to move from p along μ to the correspondingly numbered vertex r of $\zeta(\mu)$ in such fashion that the distance along μ from p to p_t changes at a rate equal to the distance along μ between p and r. The curve μ is thereby deformed into $\zeta(\mu)$. Moreover $F(\zeta) \leqq F(\mu)$. The deformation $\theta(\eta, m)$ followed by the deformation Z taken with a sufficiently small neighborhood U of η will satisfy (β).

We continue with a proof of (γ).

(γ). *A j-cap u associated with* η *rel S is a j-cap associated with* η *rel* Ω.

Since u is associated with η rel S, u is c-homologous on S to a j-cap on an arbitrarily small neighborhood of η, in particular to a j-cap v on U. If v were c-homologous to 0 on Ω, v would be c-homologous to 0 on U, by virtue of Theorem 10.1 (b). Let v^* be the final image of v under the deformation D of (β). It is clear that v^* is c-homologous to 0 on S if v is c-homologous to O on U. But v and v^* are c-homologous on S if U is sufficiently small as one sees with the aid of a deformation similar to Z operating on the formal deformation chain defined by v under D holding v and v^* fast. We infer that u is c-homologous to 0. From this contradiction we infer the truth of (γ).

We come finally to a proof of (δ).

(δ). *Any j-cap u associated with* η *rel* Ω *is c-homologous on* Ω *to a j-cap associated with* η *rel S.*

The k-cap u is c-homologous on Ω to a k-cap on the neighborhood U. It follows from (β) that u is c-homologous on Ω to a j-cap on S, and the proof of (δ) is complete.

We turn to the proof of the theorem. It follows from (γ) and (δ) that a maximal group of j-caps associated with η rel S is a maximal group of j-caps associated with η rel Ω. It then follows from (α) that the dimension of such a group is δ_k^j.

The proof of Theorem 15.2 is complete.

With the aid of this theorem and theorems of a like nature the reader can make connections with the principal results in M1.

THE INSTITUTE FOR ADVANCED STUDY.

Annals of Mathematics
Vol. 38, No. 2, April, 1937

A NOTE ON A PREVIOUS PAPER "NEW FOUNDATIONS OF PROJECTIVE AND AFFINE GEOMETRY"[1]

By Franz Alt and Karl Menger

(Received January 25, 1937)

1. This first paragraph is an erratum to the paper cited above. In §2, postulate IV (p. 458, referring to the symbol $\subseteq$ which is only afterwards introduced, has to be replaced by the two propositions (a) and (a′) of p. 458. Nothing else has to be changed in §2 since it is these hypotheses which are actually used. In §3 the remark of p. 461 that law 5 is equivalent to IV must be replaced then, by the statement that law 5, though stronger than IVa and IVa′, still holds in projective as well as in affine spaces.

2. The next two paragraphs are remarks by F. Alt on the paper cited above. It is obvious that law 5 implies IVa and IVa′. It may be shown by an example that the converse is not true. We consider a class which besides V and 9 points 1, 2, 3, 4, 1′, 2′, 3′, 4′, 0, contains 12 "lines" consisting of the points (0, 1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4), (0, 1′, 2′), (1′, 3′), (1′, 4′), (2′, 3′), (2′, 4′), (3′, 4′), and 8 "planes" consisting of the points (0, 1, 2, 3), (0, 1, 2, 4), (1, 3, 4), (2, 3, 4), (0, 1′, 2′, 3′), (0, 1′, 2′, 4′), (1′, 3′, 4′), (2′, 3′, 4′) and U. For two elements A, B of this class we define $A + B$ as the element of least dimension containing all the points of A and of B (e.g. we have $1 + 1' = 1 + 2' = U$), and $A \cdot B$ as the element consisting of the points contained as well in A as in B. Then the class satisfies Postulates I, II, III, IVa, IVa′, V, while law 5 does not hold. This is seen for $A = 0, B = (0, 1, 2), C = (0, 1, 2, 3)$.

Another remark may be added regarding this example: If we consider the subclass consisting of all parts of the plane (0, 1, 2, 3), i.e. V, 0, 1, 2, 3, (0, 1, 2), (1, 3), (2, 3), (0, 1, 2, 3) then we notice that this sub-class does not satisfy Postulate IVa. This is seen for $A = (0, 1, 2)$, $B = 0$. On the contrary, classes which satisfy laws 1–6 are "hereditary", i.e. for each element A the class of all parts of A satisfies laws 1–6 itself, A playing the rôle of U.

Vienna

[1] Annals of Mathematics, vol. 37, (1936), pp. 456–482.

ANNALS OF MATHEMATICS
Vol. 38, No. 2, April, 1937

p-ADIC ANALYSIS AND ELEMENTARY NUMBER THEORY

BY MAX ZORN*

(Received May 1, 1936)

Introduction. In 1933 I. Schur[1] communicated a far reaching generalization of Fermat's theorem in elementary number theory. He introduced as the *derivative* a_n' of a sequence a_n with respect to the number p the sequence

[0, 1] $$a_n' = \frac{a_{n+1} - a_n}{p^{n+1}}.$$

If p is a rational prime, a an integer, and $a_n = a^{p^n}$, then by Fermat's theorem

(0, 2) $$a_n' = \frac{a^{p^{n+1}} - a^{p^n}}{p^{n+1}}$$

is *integral.*

Schur proved the generalization: *Not only the first derivative* (0, 2), *but also the following derivatives up to the* $(p - 1)^{st}$ *of the sequence* a^{p^n} *are integral, if* a *is prime to* p.

His proofs however are not very simple. It is easy to see that the second derivative is integral (cf. (4.9)); but my first attempts to generalize the procedure in this case were not successful.

Later, Mr. Rothgiesser gave an inductive proof for all the statements in Schur's paper.[2]

The present paper is concerned with the discussion of Schur's theory.

First of all I call attention to the fact that this theory is valid in the p-adic field belonging to the prime number p and that the "p-adic" theorem implies again the "rational" one.

From the definition of limit in the p-adic sense we see that the sequence a_n is certainly convergent if the derived sequence a_n' is integral. Hence the Fermat-Schur theorem establishes a series of functions of the p-adic number a, namely the limits

[0, 3] $$D^{(i)}(a) = \lim_{n\to\infty} (a^{p^n})^{(i)}.$$

It turns out that it is much more convenient to study the related sequence

[0, 4] $$L_n(a) = \frac{a^{(p-1)p^n} - 1}{p^{n+1}}$$

and its derivatives. It is easy to return to the original sequence a^{p^n} (cf. §8).

---

* Sterling Fellow. Presented to the A. M. S. in October 1935.

[1] Sitzungsberichte der Preuss. Ak. d. Wiss., Math. N. Kl., 1933, p. 145.

[2] Letter to Prof. Schur., Sept. 9, 1933.

We show that the limits of the derivatives of [0, 4] exist and are given by the formula

$$(0,5) \qquad \lim_{n\to\infty} L_n^{(r)} = \frac{(p^r-1)(p^{r-1}-1)\cdots(p-1)}{(r+1)!}\left(\frac{\log a^{p-1}}{p}\right)^{r+1}.$$

Here the function $\log a^{p-1}$ may be defined either by the power series for the log or by (0, 5) itself for $r = 0$:

$$(0,6) \qquad \lim_{n\to\infty} L_n(a) = \frac{\log a^{p-1}}{p}.$$

(0, 5) has its counterpart in real analysis:

$$(0,7) \qquad \left[\frac{d^r}{dx^r}(a^x)\right]_{x=0} = (\log a)^r$$

which for $r = 1$ contains

$$(0,8) \qquad \lim_{n\to\infty} \frac{a^{\frac{1}{n}}-1}{\frac{1}{n}} = \log a.$$

But the integrity of a_n' means more than simple p-adic convergence of a_n; it gives information about the approach to the limit, or in other words congruence relations for the a_n. We obtain a system of congruences, which generalize Schur's results. I mention here only the

THEOREM. *If p is odd and $x \equiv 1$ (mod. p), and*

$$[0,9] \qquad X_n = \frac{x^p-1}{p^{n+1}},$$

then

$$(0,10) \qquad X_n^{(r)} \equiv \frac{(p^r-1)(p^{r-1}-1)\cdots(p-1)}{(r+1)!}X_n^{r+1} \pmod{p^n}$$

for $r < p$; for $r < p-2$ the congruences are valid mod. p^{n+1}.

The arrangement given in this paper is *not* intended to be a systematic simplification of Schur's theory. Even the new congruence (0, 10) would be only a formal justification, since doubtless there will be a satisfactory proof within the framework of the usual congruence arithmetics.

The interpretation in p-adic analysis shows the possibility of vast generalizations and of a direct way to *find* them. The general principle of this method would be:

Ordinary analysis has amassed a great stock of identities between power series. Many of these are valid in p-adic analysis too. But here an identity between power series yields congruences between partial sums.[3]

[3] Cf. Eisenstein, Crelle, vol. 39, page 231.

In our case the basic identity is the addition theorem of the logarithm, hence, in the last instance, Abel's theorem.

The next step in this direction would be an investigation of elliptic functions, where I expect relationships with the theory of complex multiplication.

Finally, with a certain hesitation, I shall point out another reason for the desirability of a closer study of special p-adic functions.

Various facts in the theory of algebraic numbers indicate the possibility of an "analytic continuation" of p-adic numbers. If such a "systematic identification of (algebraic and transcendental) p-adic numbers in different prime spots" exists at all, then probably the values of an analytic function, e.g.

$$\log a = \frac{1}{1-p}\sum_1^\infty \frac{(1-a^p)^n}{n}$$

for a *fixed* a would form a "monogenic system of p-adic numbers."

In this way the study of special p-adic functions might furnish material for later structural analysis.

Whether or not such monogenic systems of p-adic numbers do permit an intrinsic description is certainly an important problem. The relation to number theory proper is suggested by the formula

$$\text{Legendre symbol } \left(\frac{a}{p}\right) = \lim_{n\to\infty} a^{\frac{p-1}{2}p^n}.$$

The law of reciprocity, which describes the dependence of this function on p, has long been considered as an analogue or as a manifestation of an analogue of the Cauchy integral theorem.[4, 5]

1. The results of I. Schur.

[1, 1] Definition: Given a sequence of numbers $a_0, a_1, \cdots a_n \cdots$ and a number $p \neq 0$. Then the (Schur-)derivative Δa_n of this sequence is another sequence

$$a_n' = \Delta a_n = \frac{a_{n+1} - a_n}{p^{n+1}}.$$

Higher derivatives are defined as usual:

$$\Delta^\nu a_n = a_n^{(\nu)} = \Delta(a_n^{(\nu-1)}) = \Delta(\Delta^{\nu-1}a_n).$$

(1, 2) Fermat's Theorem. *If a is an integer and p a prime, then*

$$\Delta a^{p^n} = \frac{a^{p^{n+1}} - a^{p^n}}{p^{n+1}}$$

consists of integral numbers.

---

[4] Cf. Hilbert, *Mathematische Probleme*, Coll. Pap. vol. III., page 313.

[5] A. Weil located several years ago the analogue more precisely. It turns out that Riemann's bilinear period relation is the exact correspondent of the law of reciprocity.

We show that the limits of the derivatives of [0, 4] exist and are given by the formula

$$\lim_{n\to\infty} L_n^{(r)} = \frac{(p^r - 1)(p^{r-1} - 1)\cdots(p-1)}{(r+1)!}\left(\frac{\log a^{p-1}}{p}\right)^{r+1}. \tag{0, 5}$$

Here the function $\log a^{p-1}$ may be defined either by the power series for the log or by (0, 5) itself for $r = 0$:

$$\lim_{n\to\infty} L_n(a) = \frac{\log a^{p-1}}{p}. \tag{0, 6}$$

(0, 5) has its counterpart in real analysis:

$$\left[\frac{d^r}{dx^r}(a^x)\right]_{x=0} = (\log a)^r \tag{0, 7}$$

which for $r = 1$ contains

$$\lim_{n\to\infty} \frac{a^{\frac{1}{n}} - 1}{\frac{1}{n}} = \log a. \tag{0, 8}$$

But the integrity of a_n' means more than simple p-adic convergence of a_n; it gives information about the approach to the limit, or in other words congruence relations for the a_n. We obtain a system of congruences, which generalize Schur's results. I mention here only the

THEOREM. *If p is odd and $x \equiv 1$ (mod. p), and*

$$X_n = \frac{x^p - 1}{p^{n+1}}, \tag{[0, 9]}$$

then

$$X_n^{(r)} \equiv \frac{(p^r - 1)(p^{r-1} - 1)\cdots(p-1)}{(r+1)!} X_n^{r+1} \pmod{p^n} \tag{0, 10}$$

for $r < p$; for $r < p - 2$ the congruences are valid mod. p^{n+1}.

The arrangement given in this paper is *not* intended to be a systematic simplification of Schur's theory. Even the new congruence (0, 10) would be only a formal justification, since doubtless there will be a satisfactory proof within the framework of the usual congruence arithmetics.

The interpretation in p-adic analysis shows the possibility of vast generalizations and of a direct way to *find* them. The general principle of this method would be:

Ordinary analysis has amassed a great stock of identities between power series. Many of these are valid in p-adic analysis too. But here an identity between power series yields congruences between partial sums.[3]

[3] Cf. Eisenstein, Crelle, vol. 39, page 231.

In our case the basic identity is the addition theorem of the logarithm, hence, in the last instance, Abel's theorem.

The next step in this direction would be an investigation of elliptic functions, where I expect relationships with the theory of complex multiplication.

Finally, with a certain hesitation, I shall point out another reason for the desirability of a closer study of special p-adic functions.

Various facts in the theory of algebraic numbers indicate the possibility of an "analytic continuation" of p-adic numbers. If such a "systematic identification of (algebraic and transcendental) p-adic numbers in different prime spots" exists at all, then probably the values of an analytic function, e.g.

$$\log a = \frac{1}{1-p}\sum_{1}^{\infty}\frac{(1-a^p)^n}{n}$$

for a *fixed* a would form a "monogenic system of p-adic numbers."

In this way the study of special p-adic functions might furnish material for later structural analysis.

Whether or not such monogenic systems of p-adic numbers do permit an intrinsic description is certainly an important problem. The relation to number theory proper is suggested by the formula

$$\text{Legendre symbol} \left(\frac{a}{p}\right) = \lim_{n\to\infty} a^{\frac{p-1}{2}p^n}.$$

The law of reciprocity, which describes the dependence of this function on p, has long been considered as an analogue or as a manifestation of an analogue of the Cauchy integral theorem.[4, 5]

1. The results of I. Schur.

[1, 1] DEFINITION: Given a sequence of numbers $a_0, a_1, \cdots a_n \cdots$ and a number $p \neq 0$. Then the (Schur-)derivative Δa_n of this sequence is another sequence

$$a_n' = \Delta a_n = \frac{a_{n+1} - a_n}{p^{n+1}}.$$

Higher derivatives are defined as usual:

$$\Delta^\nu a_n = a_n^{(\nu)} = \Delta(a_n^{(\nu-1)}) = \Delta(\Delta^{\nu-1}a_n).$$

(1, 2) FERMAT'S THEOREM. *If a is an integer and p a prime, then*

$$\Delta a^{p^n} = \frac{a^{p^{n+1}} - a^{p^n}}{p^{n+1}}$$

consists of integral numbers.

[4] Cf. Hilbert, *Mathematische Probleme*, Coll. Pap. vol. III., page 313.

[5] A. Weil located several years ago the analogue more precisely. It turns out that Riemann's bilinear period relation is the exact correspondent of the law of reciprocity.

We shall say "Δa^{p^n} is integral."

This interpretation of Fermat's theorem is justified by the

(1, 3) THEOREM OF SCHUR. *If a is an integer and p a prime not dividing a, then not only Δa^{p^n}, but also the following derivatives*

$$\Delta^2 a^{p^n}, \Delta^3 a^{p^n} \cdots \Delta^{p-1} a^{p^n}$$

up to the $(p - 1)^{st}$ are integral.

From now on we shall confine ourselves (in the whole paper) to odd primes, for the sake of simplicity. This evidently does not affect (1, 3).

(1, 4) THEOREM (Schur): *If under the assumptions of* (1, 3)

$$\frac{a^{p-1} - 1}{p} \equiv 0 \text{ (mod. } p)$$

then every derivative $\Delta^r a^{p^n}$ is integral. If

$$\frac{a^{p-1} - 1}{p} \not\equiv 0 \text{ (mod. } p)$$

then (any number of) $\Delta^p a^{p^n}$ has exactly the denominator p.

We suppose the reader familiar with the elementary properties of p-adic numbers and state without proof

(1, 5) *The theorems* (1, 2)–(1, 4) *are still true if we allow a to be a p-adic number, the terms "integral," "divide" and "prime to" given their p-adic meaning.*

Moreover we see

(1, 6) *If* (1, 2)–(1, 4) *are true in the p-adic sense, then the original statements of Schur are true.*

This is due to the fact that for rational integers a the denominators of the derivatives are trivially powers of p.

2. Preliminary p-adic interpretation.

(2, 1) *If Δa_n is integral then the sequence a_n is convergent and the limit is congruent to a_0 (mod. p).*

Of course a_n would already be convergent if Δa_n had bounded denominators, and this would still be a very particular and strong form of convergence. (Cf. the relationship between continuity and differentiability of functions).

Fermat's theorem (1, 2) consequently implies the convergence of a^{p^n} for each p-adic integer a. We define

[2, 2] $$\chi(a) = \lim_{n \to \infty} a^{p^n}$$

and state

(2, 3) *The function $\chi(a)$, defined for any p-adic integer a,*

α) is multiplicative, $\chi(a)\chi(b) = \chi(ab)$

β) satisfies $\chi(a)^{p-1} = 1$, if $(a, p) = 1$,

γ) satisfies $\chi(a) = 0$, if $a \equiv 0$ (mod. p),

δ) makes $\chi(a) = \chi(b)$, only if $a \equiv b$ (mod. p).

Incidentally we have the formula

$$\chi\left(a^{\frac{p-1}{2}}\right) = \lim a^{\frac{p-1}{2}\cdot p^n} = \text{Legendre symbol } \left(\frac{a}{p}\right).$$

Up to this point we have been restating known facts; the following interpretation concerning the second derivative is new and as matter of fact was the point of departure for the further investigations.

Instead of studying Δa^{p^n} directly we consider

[2, 4] $$L_n(a) = \frac{a^{(p-1)p^n} - 1}{p^{n+1}}.$$

(2, 5) *The formula*

$$\Delta a^{p^n} = a^{p^n} \cdot L_n(a)$$

shows the connection; in particular $L_n(a)$ has the same denominator as Δa^{p^n}, if a is prime to p.

Now we know from (1, 2) that in case $(a, p) = 1$, Δa^{p^n} has an integral derivative, hence from (2, 1) that it converges; a^{p^n} converges on account of (2, 3)β) to a $(p-1)^{st}$ root of unity, in particular not to 0; consequently

(2, 6) *$L_n(a)$ converges and thus defines for $(a, p) = 1$ a function $L(a)$:*

[2, 7] $$L(a) = \lim_{n\to\infty} \frac{a^{(p-1)p^n} - 1}{p^{n+1}}.$$

This function is except for a factor the p-adic logarithm of a.

(2, 8) $$L(ab) = L(a) + L(b).$$

The proof is easy: if $a \equiv b \equiv 1 \pmod{m}$, then

$$\frac{ab-1}{m} \equiv \frac{a-1}{m} + \frac{b-1}{m} \pmod{m};$$

now we have

$$a^{(p-1)p^n} \equiv b^{(p-1)p^n} \equiv 1 \pmod{p^{n+1}},$$

hence

$$\frac{a^{(p-1)p^n} - 1}{p^{n+1}} + \frac{b^{(p-1)p^n} - 1}{p^{n+1}} \equiv \frac{(ab)^{(p-1)p^n} - 1}{p^{n+1}} \pmod{p^{n+1}}$$

or

$$L_n(a) + L_n(b) \equiv L_n(ab) \pmod{p^{n+1}};$$

$n \to \infty$ yields the theorem.

Two procedures are now possible. Either one develops the properties of $L(a)$ and the analogue of the exponential function from this definition ([2, 7]). This would correspond in real analysis to basing the theory on (0, 8) and

$$\exp x = \lim_{n\to\infty}\left(1 + \frac{x}{n}\right)^n.$$

Or the one introduces the functions log and exp by means of power series; the standard functional relations are easily derived therefrom. Based on these *and* the power series developments the properties of not only $L(a)$ and $L_n(a)$, but also of $\Delta^r L_n(a)$ and finally even the properties of $\Delta^r a^{p^n}$ can be obtained by almost straightforward computation.

3. The power series for log **and** exp. In this paragraph we shall collect the simplest standard properties of the exponential and logarithmic functions in the p-adic case, based as usual[6] on the power series

$$[3,1] \qquad \exp z = \sum_{\nu=0}^{\infty} \frac{z^\nu}{\nu!}$$

$$[3,2] \qquad \log x = -\sum_{\nu=1}^{\infty} \frac{(1-x)^\nu}{\nu}.$$

(3, 3) THEOREM: [3, 1] *has a sense* (*i.e. the series converges*) *for* $z \equiv 0 \pmod{p}$;

[3, 2] *has a sense for* $x \equiv 1 \pmod{p}$.

(3, 4) THEOREM: α) $\exp(z_1 + z_2) = (\exp z_1)(\exp z_2)$

β) $\exp z \equiv 1 \pmod{p}$.

(We use the letter z for numbers $\equiv 0 \pmod{p}$ and letters x for numbers $\equiv 1 \pmod{p}$.)

(3, 5) THEOREM: α) $\log(x_1 x_2) = \log x_1 + \log x_2$

β) $\log x \equiv 0 \pmod{p}$.

(3, 6) THEOREM: *If p is odd and*

$$x \equiv 1 \pmod{p^\rho}, \qquad x \not\equiv 1 \pmod{p^{\rho+1}},$$

then

$$\log x \equiv x - 1 \pmod{p^{2\rho}}$$
$$\not\equiv x - 1 \pmod{p^{2\rho+1}}.$$

Indeed

$$\log x - (x-1) = -(1-x)^2 \sum_{\nu=2}^{\infty} \frac{(1-x)^{\nu-2}}{\nu};$$

$\sum_{\nu=2}^{\infty}$ is congruent (mod. p) to its first member, $\frac{1}{2}$; hence the difference is exactly divisible by $p^{2\rho}$.

[6] Cf. Hensel, *Zahlentheorie*, p. 131 et seq.

Finally since $\log x \equiv 0 \pmod{p}$, $\exp z \equiv 1 \pmod{p}$, the expressions exp ($\log x$) and log ($\exp z$) have a sense, and satisfy as in real analysis the

(3, 7) THEOREM:

$$\exp(\log x) = x$$
$$\log(\exp z) = z.$$

This implies for instance that $\log x = 0$ only for $x = 1$; this follows also from (3, 6).

The statements of the following paragraphs will be proved on the basis of [3, 1] – (3, 7) alone, whereas in §§1, 2 we assumed for purposes of analysis and explanation some of Schur's results.

4. **Computation of $L(a)$.** If $x \equiv 1 \pmod{p}$, then $x^{p^n} \equiv 1 \pmod{p^{n+1}}$. Hence by the first part of (3, 6) we get

$$\log x^{p^n} \equiv x^{p^n} - 1 \pmod{p^{2n+2}}. \tag{4, 1}$$

If we use the functional equation of the log, we obtain

$$p^n \log x \equiv x^{p^n} - 1 \pmod{p^{2n+2}}. \tag{4, 2}$$

We may divide this (i.e. both sides and the modulus) by p^{n+1}; we retain integral numbers, since $\log x \equiv 0 \pmod{p}$, and $x^{p^n} - 1 \equiv 0 \pmod{p^{n+1}}$, hence:

(4, 3) THEOREM:
$$\frac{\log x}{p} \equiv \frac{x^{p^n} - 1}{p^{n+1}} \pmod{p^{n+1}}.$$

Now in (4, 3) the lefthand member is independent of n; consequently we may assert

(4, 4) THEOREM:
$$\frac{\log x}{p} = \lim_{n\to\infty} \frac{x^{p^n} - 1}{p^{n+1}}.$$

Note that in practically the same fashion one could derive (2, 9) from the definition [3, 2], in the real field at least, for $|x - 1| < 1$.

Furthermore (4, 3) includes the congruence

$$\frac{x^{p^{n+1}} - 1}{p^{n+2}} \equiv \frac{x^{p^n} - 1}{p^{n+1}} \pmod{p^{n+1}} \tag{4, 5}$$

or, returning to the Δ-terminology

(4, 6) THEOREM:
$$\Delta\left(\frac{x^{p^n} - 1}{p^{n+1}}\right) \textit{ is integral.}$$

Let $(a, p) = 1$, $a^{p-1} = x$, therefore $x \equiv 1 \pmod{p}$; then from the definition [2, 4] of $L_n(a)$ follows the

(4, 7) THEOREM. *$\Delta L_n(a)$ is integral.*

From (4, 4) arises (cf. [2, 7])

(4, 8) THEOREM: $$L(a) = \lim_{n\to\infty} L_n(a) = \frac{\log a^{p-1}}{p}.$$

There is also a more elementary proof for (4, 6) and (4, 7), based on the formula

$$x^{p^{n+1}} = (x^{p^n})^p = (1 + p^{n+1} X_n)^p \tag{4, 9}$$

where X_n is defined by

$$X_n = \frac{x^{p^n} - 1}{p^{n+1}}. \qquad [4, 10]$$

Applying the binomial theorem one obtains

$$x^{p^{n+1}} - 1 = \sum_{\nu=1}^{p} \binom{p}{\nu} p^{(n+1)\nu} X_n^\nu \tag{4, 11}$$

and

$$X_{n+1} = X_n + \sum_{\nu=2}^{p} \binom{p}{\nu} p^{(n+1)(\nu-2)-1} X_n^\nu;$$

hence for ΔX_n
(4, 12) THEOREM:

$$\Delta X_n = \sum_{\nu=2}^{p} \binom{p}{\nu} p^{(n+1)(\nu-2)-1} X_n^\nu.$$

The members of the $\sum_{\nu=2}^{p}$ are integral, since in the only dubious case $\nu = 2$, $\binom{p}{\nu} \equiv 0$ (mod. p) by virtue of $p > 2$. This proves (4, 6) and (4, 7).

But (4, 12) carries more information; it yields for $p > 3$ a simple congruence for ΔX_n; the identity (4, 12) in the form

$$\Delta X_n = \binom{p}{2} \frac{1}{p} X_n^2 + p^{n+1} \sum_{\nu=3}^{p} \binom{p}{\nu} p^{(n+1)(\nu-3)-1} X_n^\nu \tag{4, 13}$$

shows for $p > 3$ the

(4, 14) THEOREM: $$\Delta X_n \equiv \frac{p-1}{2} X_n^2 \text{ (mod. } p^{n+1}).$$

This is a considerable refinement of (4, 6). Indeed it contains the fact that the difference $\Delta X_{n+1} - \Delta X_n$ is divisible by p^{n+1}, hence—if $p > 3$—
(4, 15) THEOREM. *$\Delta^2 X_n$ is integral.*

It is possible to proceed further in this direction; this would mean an extension of Rothgiesser's proof to the congruences of type (4, 14). I hope to be able to discuss this on another occasion.

In the following paragraph we shall give a p-adic development for X_n, which is particularly suited to the Schur-derivation.

5. **p-adic representation of X_n and $\Delta^r X_n$.** We shall now develop X_n in a power series with respect to $(\log x)/p$; the Schur derivatives may then be formed term by term, which yields a similar development for all $\Delta^r X_n$.

If we want to return from equations in p-adic analysis to congruences mod. p^n, we have only to break off the power series concerned. If one prefers to avoid the transcendental function log, one has to replace $(\log x)/p$, (mod. p^{n+1}) by X_n.

$$(5,1)\quad \begin{aligned} X_n &= \frac{x^{p^n}-1}{p^{n+1}} = \frac{1}{p^{n+1}}\{\exp(\log x^{p^n}) - 1\} \\ &= \frac{1}{p^{n+1}} \sum_{\nu=1}^{\infty} \frac{(p^n \log x)^\nu}{\nu!} = \sum_{\nu=1}^{\infty} \frac{p^{n\nu-n-1}}{\nu!} (\log x)^\nu; \end{aligned}$$

this we shall mainly use in the form

(5, 2) THEOREM:

$$X_n = \sum_{\nu=1}^{\infty} p^{(n+1)(\nu-1)} \frac{\left(\frac{\log x}{p}\right)^\nu}{\nu!}.$$

In order to get ΔX_n and $\Delta^r X_n$ we have only to compute

$$\Delta^r p^{(n+1)(\nu-1)}.$$

The formula

$$(5,3)\quad \Delta^r p^{(n+1)(\nu-1)} = p^{(\nu-r-1)(n+1)}(p^{\nu-r} - 1)(p^{\nu-(r-1)} - 1) \cdots (p^{\nu-1} - 1)$$

is easily verified by induction.

Here the first factor $p^{(\nu-r-1)(n+1)}$ is fractional whenever $\nu < r + 1$; but in this case one of the remaining factors vanishes, namely $p^{\nu-(r-(\nu-r))}$:

(5, 4) THEOREM. *The first r members of* (5, 3), *considered as a sequence in ν (i.e. for $\nu = 1, \cdots r$) are zero. The others are exactly divisible by $p^{(\nu-r-1)(n+1)}$.*

Hence in the following formula for

$$\Delta^r X_n = \sum_{\nu=1}^{\infty} (\Delta^r p^{(n+1)(\nu-1)}) \frac{\left(\frac{\log x}{p}\right)^\nu}{\nu!}$$

we could make the summation begin from $\nu = r + 1$:

(5, 5) THEOREM:

$$\Delta^r\left(\frac{x^{p^n}-1}{p^{n+1}}\right) = \Delta^r X_n = \sum_{\nu=1}^{\infty} p^{(\nu-r-1)(n+1)} \frac{(p^{\nu-r}-1)\cdots(p^{\nu-1}-1)}{\nu!}\left(\frac{\log x}{p}\right)^\nu.$$

This formula (5, 5) is the key for all further statements about the $\Delta^r X_n$.

6. Properties of $\Delta^r X_n$. We are of course interested in the denominators of the $\Delta^r X_n$, but denominators will arise only from the factors $1/\nu!$ Now $p^\nu/\nu!$ is always integral. Consequently, if

$$\log x \equiv 0 \text{ (mod. } p^2), \quad \text{or} \quad \frac{\log x}{p} \equiv 0 \text{ (mod. } p),$$

we are sure that $\frac{1}{\nu!}\left(\frac{\log x}{p}\right)^\nu$ is integral. Since in this case every term of $\sum_{\nu=1}^{\infty}$ in (5, 5) is integral, $\Delta^r X_n$ itself turns out to be integral. Now we know from (3, 6) that this favourable situation arises if and only if $x \equiv 1$ (mod. p^2):

(6, 1) THEOREM. *If $x \equiv 1$ (mod. p^2), then all $\Delta^r X_n$ are integral.*

It is further important that the first non-vanishing term in (5, 5) does not depend on n. We may write

$$(6,2)\qquad \begin{aligned} \Delta^r X_n = {} & \frac{(p-1)\cdots(p^r-1)}{(r+1)!}\left(\frac{\log x}{p}\right)^{r+1} \\ & + \sum_{\nu=r+2}^{\infty} p^{(\nu-r-1)(n+1)} \frac{(p^{\nu-r}-1)\cdots(p^{\nu-1}-1)}{\nu!}\left(\frac{\log x}{p}\right)^{\nu}. \end{aligned}$$

From this we will conclude that $\Delta^r X_n$ is congruent to the first term (which is, in general, fractional) modulo any power p^M of p, provided that the index n is greater than a number $N(r, M)$ which depends on the order of derivation r and on the closeness p^M of approximation wanted.

We determine N in such a way that, given r and M

$$\frac{p^{(\nu-r-1)(n+1)}}{\nu!} \equiv 0 \text{ (mod. } p^M)$$

for $\nu - r - 1 > 0$.

Since $p^\nu/\nu!$ is integral, it is sufficient to have

$$p^{(\nu-r-1)(n+1)-\nu} \equiv 0 \text{ (mod. } p^M)$$

or

$$(\nu - r - 1)(n + 1) - \nu - M \geqq 0$$

for

$$\nu - r - 1 > 0, \qquad n > N.$$

A possible determination is $N = M + r + 1$; consequently:

(6, 3) THEOREM. *For any $M \geqq 0$ and r we have*

$$\Delta^r X_n \equiv \frac{(p-1)\cdots(p^r-1)}{(r+1)!}\left(\frac{\log x}{p}\right)^{r+1} \text{ (mod. } p^M)$$

provided that $n > M + r$.

This statement, obtained by a very rough estimate, yields nevertheless a considerable amount of information, e.g.

(6, 4) THEOREM. *Any sequence $\Delta^r X_n$ is convergent, and for the limit we have the equation*

$$\lim_{n\to\infty} \Delta^r X_n = \frac{(p-1)\cdots(p^r-1)}{(r+1)!}\left(\frac{\log x}{p}\right)^{r+1}.$$

Also

(6, 5) THEOREM. *In case $x \not\equiv 1 \pmod{p^2}$ the denominator of $\Delta^r X_n$ for $n > r$ is exactly the denominator of $1/(r+1)!$. In particular for $r < p-1, \Delta^r X_n$ is integral, and for $r = p-1$ the exact denominator is p.*

As a matter of fact the latter part of (6, 5) concerning $r < p - 1$ and $r = p - 1$ holds without restriction on n:

(6, 6) MAIN THEOREM. *Let* $\dfrac{x^{p^n}-1}{p^{n+1}} = X_n$, $x \equiv 1$ (mod. p), $p > 2$.

Then

α) *if* $x \equiv 1$ (mod. p^2), *every $\Delta^r X_n$ is integral;*

β) *if* $x \not\equiv 1$ (mod. p^2), *$\Delta^{p-1}X_n$ has denominator p;*

γ) *$\Delta X_n, \cdots, \Delta^{p-2}X_n$ are always integral.*

α) has already been proved (v. (6, 1)); the proof for β) and γ) consists in establishing the integrity of the members in (cf. (6, 2)):

$$\sum_{\nu=r+2}^{\infty} \frac{p^{(\nu-r-1)(n+1)}}{\nu!}(p^{\nu-r}-1)\cdots(p^{\nu-1}-1)\left(\frac{\log x}{p}\right)^{\nu}. \tag{6, 7}$$

We prove now the

(6, 8) THEOREM. *$p^{(\nu-r-1)(n+1)}/\nu!$ is integral for $\nu - r - 1 > 0, r < p$.*

If $\nu = u_0 + u_1 p + \cdots + u_e p^e$, then $\nu!$ contains exactly $p^{u_1+\cdots+u_e}$; hence we have to confirm

$$(\nu - r - 1)(n + 1) - (u_1 + \cdots + u_e) \geqq 0.$$

Apparently it is sufficient to prove (6, 8) for $n = 0$:

$$(\nu - r - 1) - (u_1 + \cdots u_e) \geqq 0$$

or

$$u_0 + u_1(p-1) + \cdots + u_e(p^e - 1) \geqq r + 1. \tag{6, 9}$$

a) For $\nu < p$ (6, 9) means $u_0 \geqq r + 1$, and this follows from $u_0 = \nu > r + 1$.
b) For $\nu = p$ (6, 9) means $p - 1 \geqq r + 1$; this follows again from $p = \nu > r + 1$.
c) For $\nu > p$ the left-hand member of (6, 9) is at least p, and this is $\geqq r + 1$ by virtue of the restriction on r.

Finally we obtain a theorem about any a which is prime to p by putting $x = a^{p-1}$.

(6, 10) THEOREM: *Let* $p > 2, \quad (a, p) = 1, \quad L_n(a) = \dfrac{a^{(p-1)p^n} - 1}{p^{n+1}}$.

Then

α) *if* $a^{p-1} \equiv 1$ (mod. p^2), *every* $\Delta^r L_n(a)$ *is integral;*
β) *if* $a^{p-1} \not\equiv 1$ (mod. p^2), $\Delta^{p-1} L_n(a)$ *has the denominator* p;
γ) $\Delta L_n(a), \cdots, \Delta^{p-2} L_n(a)$ *are always integral.*

7. **Congruences** mod. p^n **for** $\Delta^r X_n$. In order to obtain sharper results we need only scrutinize our proofs concerning the members of (6, 7) in (6, 2).

First of all we considered $p^{\nu-r-1}/\nu!$ alone, disregarding a factor $p^{(\nu-r-1)n}$ which is at least p^n. This is necessary if we do not want the modulus to depend on n. Otherwise we might have stated that $\Delta^r X_n$ is congruent to its limit (cf. (6, 4)) modulo p^n in all cases covered by (6, 6) or (6, 10).

On the other hand we did not investigate the actual p-exponent of (6, 7). We shall ask now in this direction: When is $p^{\nu-r-1}/\nu!$, which we know to be integral (for $\nu > r + 1$ and $r < p$), *not* divisible by p?

In this case we should have

$$u_0 + u_1(p-1) + \cdots + u_e(p^e - 1) = r + 1 \tag{7, 1}$$

where

$$\nu = u_0 + u_1 p + \cdots + u_e p^e > r + 1, \qquad u_e \geqq 1. \tag{7, 2}$$

$$r + 1 \leqq p. \tag{7, 3}$$

a) $\nu < p$, we have $u_0 = \nu$, $u_i = 0$ for $i \neq 0$.
Since $\nu > r + 1$, (7, 1) cannot be true.

b) If $\nu = p$, then (7, 1) gives $p - 1 = r + 1$ or $r = p - 2$, $\nu = r + 2$.
In this case, then, the first member of (6, 7) is exactly divisible by p^n, provided that the logarithmic factor is prime to p. The other members are divisible by p^{n+1}, hence the infinite sum (6, 7) is also exactly divisible by p^n.

c) If $\nu > p$ we know that

$$u_0 + u_1 p + \cdots + u_e p^e > p$$

and

$$u_0 + u_1(p-1) + \cdots + u_e(p^e - 1) = r + 1 \leqq p.$$

The second inequality can be replaced by the equality by virtue of the first; we obtain $r = p - 1$, $\nu = p + 1$. In other words $\dfrac{p^{\nu-r-1}}{\nu!}$ is not divisible by p exactly when $(r = p - 1)\ \nu = r + 2$.

Again we see that $\Delta^r X_n$ in general will be congruent to its limit exactly modulo p^n.

We collect these results in the

(7, 4) MAIN THEOREM:

α)
$$\Delta^r X_n \equiv \frac{(p^r-1)\cdots(p-1)}{(r+1)!}\left(\frac{\log x}{p}\right)^{r+1} (\text{mod. } p^{n+1})$$

for $r = 1, 2, \cdots p-3$.

β)
$$\Delta^r X_n \equiv \frac{(p^r-1)\cdots(p-1)}{(r+1)!}\left(\frac{\log x}{p}\right)^{r+1} (\text{mod. } p^{n})$$

for $r = p-2, p-1$.

γ)
$$\Delta^r X_n \not\equiv \frac{(p^r-1)\cdots(p-1)}{(r+1)!}\left(\frac{\log x}{p}\right)^{r+1} (\text{mod. } p^{n+1})$$

if $r = p-2, p-1$, *and* $\dfrac{\log x}{p} \not\equiv 0$ (mod. p).

Finally we write (7, 4) without logarithms, replacing $(\log x)/p$, modulo p^{n+1} by X_n (cf. (4, 31)).

(7, 5) THEOREM:

α)
$$\Delta^r X_n \equiv \frac{(p^r-1)\cdots(p-1)}{(r+1)!} X_n^{r+1} (\text{mod. } p^{n+1})$$

for $r = 1, 2, \cdots p-3$;

β)
$$\Delta^r X_n \equiv \frac{(p^r-1)\cdots(p-1)}{(r+1)!} X_n^{r+1} (\text{mod. } p^{n}),$$

for $r = p-2, p-1$;

γ)
$$\Delta^r X_n \not\equiv \frac{(p^r-1)\cdots(p-1)}{(r+1)!} X_n^{r+1} (\text{mod. } p^{n+1})$$

for $r = p-2, p-1$ *and* $X_1 \not\equiv 0$ (mod. p).

8. **Proof of Schur's statements about $\Delta^r a^{p^n}$.** As stated before in §2 we get information about $\Delta^r a^{p^n}$ from

(2, 5)
$$\Delta a^{p^n} = a^{p^n} \cdot L_n(a).$$

Evidently we need formulas which connect the derivative $\Delta(a_n b_n)$ of $a_n b_n$ with the derivatives Δa_n and Δb_n. We state these rules in the

(8, 1) THEOREM.

α) $\Delta(a_n + b_n) = \Delta a_n + \Delta b_n$

β) $\Delta(a_n b_n) = a_n \Delta b_n + (\Delta a_n) b_{n+1}$

γ) $\Delta(c a_n) = c \Delta a_n$.

Now let us take the derivative of (2, 5):

$$(8, 2) \qquad \begin{aligned} \Delta^2 a^{p^n} &= a^{p^n}\Delta L_n(a) + (\Delta a^{p^n})\cdot L_{n+1}(a) \\ &= a^{p^n}(\Delta L_n(a) + L_n(a)L_{n+1}(a)). \end{aligned}$$

By virtue of (6, 10) we get for $p > 2$, $(a, p) = 1$, the

(8, 3) THEOREM: *$\Delta^2 a^{p^n}$ is integral.*

It is possible to derive more precise congruences for these and higher derivatives. This will be done in another paper. Here we content ourselves with the integrity of the first $(p - 1)$ derivatives and the denominator for the p^{th} derivative. For this purpose we do not need the precise expression of the higher derivatives in terms of the $\Delta^\nu L_{n+\mu}$. We use only

$$(8, 4) \qquad \Delta^r a^{p^n} = a^{p^n}(\Delta^{r-1}L_n(a) + P)$$

where P is a polynomial with integral coefficients in the $\Delta^\nu L_{n+\mu}(a)$ with order of derivation ν less than $r - 1$.

This is easily proved by induction.

Hence if $\Delta L_n(a), \cdots, \Delta^{r-1}L_n(a)$ are integral for any n, then $\Delta^r a^{p^n}$ is integral too. Moreover, under these assumptions $\Delta^{r+1}a^{p^n}$ differs from $a^{p^n}\Delta^r L_n(a)$ only by an integral number, and thus these numbers have both the same denominator. This finishes the proof of Schur's main theorem about the $\Delta^r a^{p^n}$, since the $\Delta^\nu L_n(a)$ occurring in their representation according to (8, 4) are sufficiently known from (6, 10):

(8, 5) MAIN THEOREM OF SCHUR: *If p is a rational prime > 2, a an integer prime to p, then*

α) $\Delta a^{p^n}, \cdots, \Delta^{p-1}a^{p^n}$ are always integral;

β) if $a^{p-1} \equiv 1 \pmod{p^2}$, then all $\Delta^r a^{p^n}$ are integral;

γ) if $a^{p-1} \not\equiv 1 \pmod{p^2}$, then $\Delta^p a^{p^n}$ has denominator p.

The statement holds in the rational or in the p-adic field.

YALE UNIVERSITY.

ANNALS OF MATHEMATICS
Vol. 38, No. 2, April, 1937

COLUMN NORMAL MATRIC POLYNOMIALS

BY MERRILL M. FLOOD

(Received November 10, 1936)

The principal questions answered in this paper are:

I. What is the greatest lower bound for $d[PX]$ when P is a given matric polynomial, X is a matric polynomial of prescribed rank, and $d[PX]$ is the degree of the product PX?

II. How can the degree invariants of the column vector space of $C_r[P]$ be expressed in terms of the degree invariants of the column vector space of P, where P is a given matric polynomial and $C_r[P]$ is its r^{th} compound?

III. How can the problem of finding the matric roots of a matric polynomial P be reduced to the consideration of regular matric polynomials?

1. **Matric polynomials.** Let e^{pq}_{rs} be the matrix with p rows and q columns, all of whose elements are zero except the one in the r^{th} row and s^{th} column which is unity. The matrix $A^{pq}[\lambda] = \sum_{r=1}^{p} \sum_{s=1}^{q} A^{rs}[\lambda]\, e^{pq}_{rs}$, where $A^{rs}[\lambda]$ are polynomials whose coefficients and variable λ lie in the complex field, may be written in the form $A^{pq}[\lambda] = \sum_{k=0}^{n} A^{pq}{}_k \lambda^k$ where $A^{pq}{}_n \neq 0$. The matrix $A^{pq}[\lambda]$ is called a *matric polynomial* of *order* $[p, q]$ and *degree* n with coefficients $A^{pq}{}_k$ *over* the complex field.

It will be convenient to denote the degree of $A^{pq}[\lambda]$ by $d[A^{pq}[\lambda]]$, its rank by $p[A^{pq}[\lambda]]$, and its transpose by $A'^{pq}[\lambda]$. Furthermore, $A^{pq}[\lambda]$ will be called *regular* if $p[A^{pq}[\lambda]] = \min [p, q]$, *square* if $p = q$, and *elementary* if $[A^{pq}[\lambda]]^{-1}$ is also a matric polynomial.

2. **Column normality.** For simplicity, set $P = A^{pq}[\lambda]$, $\rho = p[A^{pq}[\lambda]]$, $m = q - \rho$, $e_k = e^{q1}_{k1}$, and $d_k[P] = d[Pe_k]$ for $(k = 1, 2, \cdots, q)$. The vector polynomials[1] Pe_k for $(k = 1, 2, \cdots, q)$ determine an integral set[1] $W[P]$ of rank ρ. The matric polynomial $A^{pq}[\lambda]$ is said to be *column normal* if and only if it satisfies the following three conditions: (i) $Pe_k = 0$ for $(k = 1, 2, \cdots, m)$, (ii) $d_{m+1}[P] \leqq d_{m+2}[P] \leqq \cdots \leqq d_q[P]$, and (iii) Pe_k for $(k = m + 1, m + 2, \cdots, q)$ form a normal basis[1] for $W[P]$. The integers $d_k[P]$ for $(k = m + 1, m + 2, \cdots, q)$ are the degree invariants[1] of $W[P]$ and are called the *column invariants* of $A^{pq}[\lambda]$. It is convenient to set $d_k[P] = 0$ for $(k = 1, 2, \cdots, m)$. (Similarly, P is said to be *row normal* if P' is column normal and its *row invariants* are the column invariants of P'.)

[1] Wedderburn, J. H. M.: *Lectures on Matrices*, p. 47–49.

It follows immediately[1] that an elementary polynomial $E^{qq}[\lambda]$ always exists such that $N^{pq}[\lambda] = A^{pq}[\lambda]E^{qq}[\lambda]$ is column normal. Of course, $E^{qq}[\lambda]$ is not unique but any polynomial such as $N^{pq}[\lambda]$ is called a *column normal form* of $A^{pq}[\lambda]$.

3. Theorem I. (1) *If* $P = P^{pq}[\lambda]$ *and* $X = X^{qr}[\lambda]$ *are matric polynomials, then* $d[PX] \geqq d_{p[X]}[P]$. (2) *If P is a matric polynomial of order* $[p, q]$ *and j and k are integers such that* $0 \leqq j \leqq q$ *and* $k \geqq j$, *then a matric polynomial X of order* $[q, k]$ *and rank j exists such that* $d[PX] = d_j[P]$.

Proof. (1) Let $N = PE$ be column normal where E is elementary and set $L = E^{-1}X$. If $d[NL] < d_{p[L]}[N]$ it follows immediately[2] that $L'e_{k1}^{q1} = 0$ for $(k = p[L], p[L] + 1, \cdots, q)$ which is clearly impossible since $p[L] = p[X]$. Hence $d[NL] \geqq d_{p[L]}[N]$ or $d[PX] \geqq d_{p[X]}[P]$. (2) Let $N = PE$ be column normal where E is elementary and set $M = \sum_{\alpha=1}^{j} e_{\alpha\alpha}^{qk}$. Obviously $X = EM$ satisfies the required conditions.

4. **Degree invariants of compounds.**[3] If P is a matric polynomial of order $[p, q]$ and E is elementary then $C_r[PE] = C_r[P]C_r[E]$. Since $C_r[E]$ is elementary it follows that the column invariants of $C_r[PE]$ are the same as those for $C_r[P]$ and so it is sufficient to consider the case in which P is a column normal matric polynomial. In fact, it will be convenient to ignore the first $q - p[P]$ columns of P and consider the case of a regular column normal matric polynomial, since this obviously involves no loss of generality.

The lemma which follows is stated without proof and might have been taken as the definition of a regular column normal matric polynomial.

Lemma A. *The regular matric polynomial* $N = N^{pq}[\lambda]$ *is column normal if and only if it satisfies the following conditions*: (i) $q = p[N]$, (ii) $d_1 \leqq d_2 \leqq \cdots \leqq d_q$, *and* (iii) *the product* $Q = ND$ *is proper* (*leading coefficient regular*), *where*

$$D = \sum_{k=1}^{q} e_{kk}^{qq} \lambda^{d-d_k}, \qquad d = d[N], \qquad \text{and} \qquad d_k = d[Ne_{k1}^{q1}].$$

Now consider the relation $C_r[Q] = C_r[N]C_r[D]$ where $C_r[D]$ is clearly a diagonal matric polynomial with powers of λ in the diagonal. Let the compound be defined so that $C_r[D] = \sum_{k=1}^{\theta} \lambda^{\beta_k} e_{kk}^{\theta\theta}$ where $\theta = \binom{q}{r}$ and $\beta_1 \geqq \beta_2 \geqq \cdots \geqq \beta_\theta$. The β_k are the various sums of the integers $d - d_1, d - d_2, \cdots, d - d_q$ taken r at a time, and in particular $\beta_\theta = \sum_{k=q-r+1}^{q} (d - d_k)$. Hence, if $Q_r = C_r[Q]/\lambda^{\beta_\theta}$ and $D_r = C_r[D]/\lambda^{\beta_\theta}$, then $Q_r = C_r[N]D_r$ and $C_r[N]$ is column normal since Q_r is proper.

Let f_k be the column invariants of $C_r[N]$ and set $\alpha_k = \beta_k - \beta_\theta$ for $(k = 1, 2,$

[2] Ibid. p. 49.
[3] Ibid. p. 64.

$\cdots, \theta)$. Then $f_k = d[C_r[N]] - \alpha_k$ but $d[Q_r] = d[C_r[N]] = d[C_r[Q]] - \beta_\theta = rd[N[- \beta_\theta$ and so $f_k = rd - \beta_k$. This proves

THEOREM II. *If P is a matric polynomial of rank p with column invariants d_k for $(k = 1, 2, \cdots, p)$, if r is an integer such that $0 < r \leqq p$, if γ_k are the various sums of the integers $d_1, d_2, \cdots, d_p$ taken r at a time and such that $\gamma_1 \leqq \gamma_2 \leqq \cdots \leqq \gamma_\theta$ where $\theta = \binom{p}{r}$, and if f_k for $(k = 1, 2, \cdots, \theta]$ are the column invariants of $C_r[P]$; then $f_k = \gamma_k$.*

5. THEOREM III. *The left zeros of a matric polynomial P of nullity n are the same as those of the regular polynomial obtained from any column normal form of P by omitting the first n columns.*

PROOF. (i) Suppose that R is a left zero of P. By the remainder theorem[4] it follows that $P = [\lambda - R]Q$ where Q is a matric polynomial. If $N = PE$ is a column normal form of P then R is a left zero of N since $N = [\lambda - R][QE]$. The first n columns of N are zero so necessarily the first n columns of QE are likewise zero. If N_0 is the polynomial obtained from N by omitting its first n columns then clearly R is a left zero of N_0 since $N_0 = [\lambda - R][QE]_0$ where $[QE]_0$ is the polynomial obtained from QE by omitting its first n columns.

(ii) Conversely, if R is a left zero of N_0 it is obviously a zero of N and hence of P so the proof is complete.

This result reduces the problem of finding left zeros of matric polynomials to the consideration of regular column normal matric polynomials.

6. **Further results.**

COROLLARY I.1. *If P is a non-singular matric polynomial whose greatest column invariant is d_q then there exists a matric polynomial R such that PR is proper and of degree d_q. Furthermore, $d[PR] \geqq d_q$ for every non-singular matric polynomial R.*

PROOF. This corollary is an immediate consequence of Theorem I and Lemma A.

COROLLARY II.1. *If N is a non-singular column normal matric polynomial of order n then the adjoint of N, with its rows in reverse order, is row normal.*

PROOF. Since $[\mathrm{adj}N]' = C_{n-1}[N]$ it follows from the proof of Theorem II that $[\mathrm{adj}N]'$ is column normal if its columns are properly rearranged. However, it is obviously only necessary to invert the order of the rows and of the columns of D_{n-1} to make $\beta_1 \geqq B_2 \geqq \cdots \geqq \beta_n$ in this case and hence only necessary to invert the columns of $[\mathrm{adj}N]'$; which has the same effect as inverting the rows of $\mathrm{adj}N$.

[4] Ibid. p. 22.

Corollary II.2. *If P is a non-singular matric polynomial of order n with column invariants d_k for $[k = 1, 2, \cdots, n]$ then the row invariants of its adjoint are $f_k = \left[\sum_{\alpha=1}^{n} d_\alpha\right] - d_{n-k+1}$ for $[k = 1, 2, \cdots, n]$.*

Proof. This follows immediately from Theorem II since $[\operatorname{adj}P]' = C_{n-1}[P]$. (It should be noted that $\sum_{\alpha=1}^{n} d_\alpha$ is the degree of the determinant of P.)

Corollary III.1. *If a matric polynomial P has a left zero R then all of its column degree invariants are greater than zero.*

Proof. If R is a left zero of P then it is a left zero of a regular column normal matric polynomial N. Hence $N = [\lambda - R]Q$ which is clearly impossible if all elements of some column of N are constants.

Princeton University.

Annals of Mathematics
Vol. 38, No. 2. April, 1937

CLASS FIELDS OF INFINITE DEGREE OVER p-ADIC NUMBER FIELDS

By Otto F. G. Schilling

(Received May 20, 1936)

In this paper we shall develop a theory of the infinite abelian algebraic extensions of p-adic number fields. We try to characterize these fields by certain groups defined in the ground field. In the case of finite abelian extensions these groups are the norm class groups[1] and our goal is to find a substitute for these. The natural approach to this is to consider the set of all finite subfields contained in the given field of infinite degree. All these subfields define norm class groups in the ground field. With the help of these groups we are able to define a system of neighbourhoods in the multiplicative group of all elements (not equal to zero) belonging to the ground field. The closure of this topologized group is the generalization of the finite class groups. It is isomorphic to the Galois group of the infinite abelian extension. To establish this fact we use the theorems of the known finite local class field theory. We may mention that the problem is slightly simplified by the remark that the infinite extensions which are considered here are all enumerable. Therefore we need only an enumerable system of neighbourhoods to describe the topology in the space mentioned above.

Before going into the details of the construction we first state some well known facts about the local class field theory and the Galois theory of enumerable infinite extensions.[2]

1. Let k be a finite p-adic number field of the most general kind, that is to say, a field which is perfect with respect to the discrete valuation B of rank one and the field of residue classes belonging to the prime ideal p of k is a finite Galois field.[3] For example, all the possible p-adic extensions of finite algebraic number fields are such fields.

Let us denote by k^* the multiplicative group of all elements in the field k that are not equal to zero. Then the factor group k^*/k^{*n} is a finite abelian group for every positive integer n.[4]

[1] Cl. Chevalley, *La théorie du symbole des restes normiques*, Crelle vol. 169 (1933).

[2] For details on infinite algebraic extensions cf. M. Moriya, *Theorie der algebraischen Zahlkörper unendlichen Grades* (MI), Journ. of the fac. of sci. Sapporo ser. I. vol. III (1935). *Galoissche Theorie der algebraischen Zahlkörper unendlichen Grades* (MII), same journal ser. I. vol. IV (1936).

[3] Such fields are, in addition to the p-adic extensions of number fields, the perfect extensions of abstract fields of functions. Cf. H. Hasse in different papers in Crelle vol. 173 (1935).

[4] For number fields e.g. K. Hensel, *Allgemeine Theorie der Kongruenzklassengruppen und ihrer Invarianten in algebraischen Zahlkörpern*, Crelle vol. 147 (1917).

In the local class field theory all abelian extensions K of degree n over k are described in a one to one way by the different subgroups of index n of the group k^*/k^{*n}.

If we denote by K^* the multiplicative group of the field K, by $N_{K/k}$ the norm of K relative to k, and by $G_{K/k}$ the Galois group of K with respect to k, then we have the isomorphism

$$G_{K/k} \cong k^*/N_{K/k}K^*.$$

This is the theorem of isomorphism.

The actual isomorphism is furnished by the norm residue symbol. For every number $a \neq 0$ of k there is defined the symbol (a, K)[5] which is an element of the galois group $G_{K/k}$. This symbol has the following properties

i) $(a, K)(b, K) = (ab, K)$,

ii) $(a, K) = 1$ if and only if $a = N_{K/k}A$ with A in K,

iii) $(a, K) \to (a, K')$ if the substitution is applied on a subfield K' of K which contains k,

iv) (a, K) runs through all elements of $G_{K/k}$ if a runs through all elements of k^*.

By property ii) of the symbol the group of all elements a of k such that $(a, K) = 1$, coincides with the norm group $N_{K/k}K^*$.

The properties i) and iv) show that the mapping of $k^*/N_{K/k}K^*$ in the group $G_{K/k}$ is an isomorphism. This theorem is called the law of reciprocity.

Let us denote the norm group, $N_{K/k}K^*$, which belongs to the abelian extension K of k by $H(K)$.

The subfields K' of K containing k correspond in a one to one way to the groups H' lying between k and $H(K) = H$, and since

$$k \subseteq K' \subseteq K \leftrightarrow k^* \supseteq H' \supseteq H,$$

we see that $H' = H(K')$. (Theorem of ordering and unicity.)

Furthermore it is shown in the local class field theory that for given subgroups H of finite index in k^* there exist abelian extensions K of k such that $H(K) = H$. (Theorem of existence.)

Now let L be an infinite normal algebraic extension of k, and let $\mathbf{G}$ denote the group of *all* automorphisms of L relative to k. If K is an arbitrary finite normal extension of k which is contained in L, let $G_{K/k} = G(K)$ again denote its Galois group with respect to k. Every automorphism $\mathbf{g}$ of $\mathbf{G}$ furnishes an automorphism $g(K)$ of $G(K)$ if $\mathbf{g}$ is applied to the elements of K. All elements $\mathbf{g}'$ of $\mathbf{G}$ which furnish the same automorphism $g(K)$ as $\mathbf{g}$ are then put in one class $U(\mathbf{g}, K)$. These classes form, for variable fields K in L, systems which have the formal properties of neighbourhoods of a topological space. The group $\mathbf{G}$ of all automorphisms of L over k is then closed with respect to the topology we have introduced.

---

[5] We write (a, K) instead of $\left(\frac{a, K/k}{p}\right)$ which is the usual notation.

The fundamental theorem of the Galois theory then states that the closed subgroups $\mathbf{S}$ of $\mathbf{G}$ correspond in a one to one way to the subfields F of L.

This most general topologization of the Galois group $\mathbf{G}$ can be slightly modified. We have already assumed that L is algebraic with respect to the p-adic number field k. This means that L can be approximated by countably many finite normal fields K_i, such that

$$k \subseteq \cdots \subseteq K_{i-1} \subseteq K_i \subseteq \cdots \subseteq L \qquad \text{and}$$

$$L \text{ is the sum of all the fields } K_i.\text{[6]}$$

The fields K_i also define neighbourhoods $U(\mathbf{g}, K_i)$; it turns out that the topology which is defined by this countable set of neighbourhoods is the same as the topology defined by all finite subfields.

Let G_i be the Galois group of the field K_i with respect to k, and let g_i denote an arbitrary element of G_i. Instead of $U(\mathbf{g}, K_i)$ we may also write $U(g_i, K_i)$ if $\mathbf{g}$ furnishes the automorphisms $g_i = g(K_i)$ in K_i. A sequence of neighbourhoods $U(g_i, K_i)$ such that

$$U(g_{i+1}, K_{i+1}) \subseteq U(g_i, K_i)$$

in the sense of the theory of sets as applied to the space $\mathbf{G}$ will be called a *fundamental sequence.* It is obvious that the elements $\mathbf{g}$ of $\mathbf{G}$ correspond uniquely to the fundamental sequences of neighbourhoods.

This fact can be expressed in a more formal way as follows. For every index i, the galois group G_{i-1} is a homomorphic map of the Galois group G_i. This homomorphism may be called ψ_i. With respect to these homomorphisms we define fundamental sequences $(g_1, g_2, \cdots, g_i, \cdots)$ of elements g_i of the groups G_i. We postulate that $g_{i-1} = g_i\psi_i$ for every index i.

The product of two sequences $(g_1, g_2, \cdots, g_i, \cdots)$ and $(g_1', g_2', \cdots, g_i', \cdots)$ will be defined as the sequence $(g_1g_1', g_2g_2', \cdots g_ig_i', \cdots)$.

Again these sequences form a topological group and one easily shows that it is isomorphic to the group $\mathbf{G}$.

An immediate consequence of the finiteness of the groups G_i is that the group $\mathbf{G}$ is a compact group.[7]

The last definition of $\mathbf{G}$ has the advantage that it can be better used for the more detailed study of the infinite fields.[8]

Finally, we observe that the formal product over all the relative degrees $(K_{i+1} : K_i) = n_i$ is a G-number $N = \lim_{i\to\infty} \prod_{j=1}^{i} n_j$ in the sense of Steinitz.[9] It depends on the special approximating sequence K_i of the field L. Later on we shall use this G-number for a special construction.

---

[6] By the sum or compositum of fields K_i we understand the smallest field K in the algebraically closed extension Ω over k which contains all the fields K_i.

[7] Cf. W. Krull, *Galoissche Theorie der unendlichen algebraischen Erweiterungen*, Math. Annalen vol. 100 (1928).

[8] Cf. MII.

[9] E. Steinitz, *Algebraische Theorie der Körper*, Crelle vol. 137 (1910).

2. Let L be an infinite abelian algebraic extension of the p-adic field k and $\{K_i\}$ an approximating sequence of finite subfields containing k. Then the norm groups $H(K_i) = N_{K_i/k}K_i^*$ for $i = 1, 2, \cdots$ form a properly descending sequence of groups in k:

$$k^* \supseteq H(K_1) \supseteq H(K_2) \supseteq \cdots \supseteq H(K_i) \supseteq \cdots.$$

The intersection of all these groups will be called $\mathbf{N}$. The group $\mathbf{N}$ is uniquely determined by the approximating sequence $\{K_i\}$.

We now form the class group $k^*/\mathbf{N}$. Its elements will be denoted by (a). This group has the property that it can be mapped on the class groups $k^*/H(K_i)$ for $i = 1, 2, \cdots$. The subgroup which corresponds to the homomorphism $k^*/\mathbf{N} \to k^*/H(K_i)$ is $H(K_i)/\mathbf{N}$, because $(k^*/\mathbf{N})/(H(K_i)/\mathbf{N}) \cong k^*/H(K_i)$. Furthermore, $k^*/H(K_i)$ is mapped homomorphically on $k^*/H(K_{i-1})$: this homomorphism will be denoted by φ_i.

These two homomorphisms suggest a topologization of the group $k^*/\mathbf{N}$ with respect to the class groups $k^*/H(K_i)$ as approximations.

In order to obtain this topology in $k^*/\mathbf{N}$ we proceed as follows. First we get a representation of $k^*/\mathbf{N}$. To do this let (a) be an arbitrary element of $k^*/\mathbf{N}$. Then (a) is congruent to a well defined residue class $\bar{a}_i$ of $k^*/\mathbf{N}$ modulo $H(K_i)/\mathbf{N}$. Let us put together all these residue classes (which can be represented as classes of $k^*/H(K_i)$ as a consequence of the isomorphism $(k^*/\mathbf{N})/(H(K_i)/\mathbf{N}) \cong k^*/H(K_i)$ in the row $(\bar{a}_1, \bar{a}_2, \cdots, \bar{a}_i, \cdots)$.

The notion of equality in $k^*/\mathbf{N}$ carries over to the representation of the elements by these rows, that is $(a) = (b)$ if and only if $(\bar{a}_1, \bar{a}_2, \cdots, \bar{a}_i, \cdots) = (\bar{b}_1, \bar{b}_2, \cdots, \bar{b}_i, \cdots)$. This is an immediate consequence of the fact that the elements (a) are classes of k^* modulo $\mathbf{N}$.

The next step is to introduce *fundamental sequences* $(\bar{c}_1, \bar{c}_2, \cdots, \bar{c}_i, \cdots) = \mathbf{c}$ of elements $\bar{c}_i$ in $k^*/H(K_i)$. A sequence $\mathbf{c}$ is called a fundamental sequence if and only if $\bar{c}_{i-1} = \bar{c}_i\varphi_i$. The definitions of equality and of product of two fundamental sequences are again obvious.

The group $\mathbf{k} = \mathbf{k}(L)$[10] of all fundamental sequences will be called the *adjoint group* of the infinite abelian algebraic extension L in the ground field k. As in the Galois theory of infinite algebraic extensions it can be shown that $\mathbf{k}$ is a compact group.[11] The group $\mathbf{k}$ contains the class group $k^*/\mathbf{N}$.

To this construction there corresponds the introduction of a system of neighbourhoods in $k^*/\mathbf{N}$. The neighbourhood $U_{i,\bar{a}_i}$ is composed of all elements (a) of $k^*/\mathbf{N}$ for which (a) is congruent to $\bar{a}_j$ of $k^*/H(K_j)$, $j = 1, 2, \cdots, i$. This means that the neighbourhood $U_{i,\bar{a}_i}$ is made up of all the elements (a) which lie in the same residue class $(a)\cdot H(K_i)/\mathbf{N}$. These systems of elements form a

---

[10] We shall see that this notation $\mathbf{k} = \mathbf{k}(L)$ is justified. The closure of $k^*/\mathbf{N}$ does not depend on the special approximation $\{K_i\}$ of L.

[11] Cf. no. 7.

denumerable set of topological neighbourhoods. It is easy to verify that all the postulates for topological neighbourhoods are fulfilled.[12]

A sequence $U_{h,\bar{a}_h}$ for $h = 1, 2, \cdots$ is called a fundamental sequence of neighbourhoods, if

$$U_{h-1,\bar{a}_{h-1}} \subseteq U_{h,\bar{a}_h}.$$

The fundamental sequences of neighbourhoods are in a one to one correspondence with the fundamental sequences $\mathbf{c}$. The group $\mathbf{k}$ is the closure of $k^*/\mathbf{N}$ with respect to this topology.

Remark: The system of neighbourhoods we just constructed could be replaced by a slightly more general system. Let K be an arbitrary finite subfield of the given field L and $H(K)$ its corresponding norm group in k. Furthermore let $\mathbf{N}'$ be the intersection of *all* $H(K)$. We then consider the class group $k^*/\mathbf{N}'$. Now all classes $(a)'$ which lie in the same residue class of $k^*/\mathbf{N}'$ modulo $H(K)/\mathbf{N}'$ constitute the neighbourhood $U_{K,\bar{a}_K}$. These sets of elements also fulfill all postulates of topological neighbourhoods.[13] The compact space $\mathbf{k}'$ with respect to this topology of $k^*/\mathbf{N}'$ is isomorphic to the space $\mathbf{k}$ which was defined with respect to a defining sequence $\{K_i\}$ of L. The reason for this fact is that each field K contained in L is a subfield of a suitably chosen field K_i of the special approximating sequence K_i. Furthermore $\mathbf{N} = \mathbf{N}'$. The denumerable system of neighbourhoods $U_{i,\bar{a}_i}$ is then equivalent to the system of all $U_{K,\bar{a}_K}$. Therefore the spaces $\mathbf{k}$ and $\mathbf{k}'$ can be considered as identical. In the following investigation we prefer the more explicit representation of $\mathbf{k}$ as the group of all fundamental sequences $\mathbf{c}$.

Theorem I. *The adjoint group* $\mathbf{k}$ *of an infinite abelian extension* L *is isomorphic to the Galois group* $\mathbf{G}$ *of* L.

Proof. We shall prove the isomorphism by writing down explicitly a mapping between $\mathbf{k}$ and $\mathbf{G}$. For this purpose we apply the theory of the norm residue symbol (a, K) which establishes in the finite case the law of reciprocity. Let us abbreviate the symbol (a, K_i) to $\chi_i(a)$. Property ii) of the finite norm residue symbol asserts that $\chi_i(a)$ depends only on the class of a modulo the norm group $H(K_i)$. Therefore we can write $\chi_i(a_i) = \chi_i(\bar{a}_i)$, where $\bar{a}_i$ denotes the class, modulo $H(K_i)$, which is uniquely determined by the element a_i.

Now let $\mathbf{a} = (\bar{a}_1, \bar{a}_2, \cdots, \bar{a}_i, \cdots)$ be an arbitrary element of the adjoint group $\mathbf{k}$. We now define a symbol $\chi(\mathbf{a})$ for all elements $\mathbf{a}$ of $\mathbf{k}$ by

$$\chi(\mathbf{a}) = (\chi_1(\bar{a}_1), \chi_2(\bar{a}_2), \cdots, \chi_i(\bar{a}_i), \cdots).$$

This symbol is multiplicative, i.e. $\chi(\mathbf{a})\cdot\chi(\mathbf{b}) = \chi(\mathbf{ab})$ for any two elements $\mathbf{a}$ and $\mathbf{b}$ of the group $\mathbf{k}$.

---

[12] $U \cap V$ is a neighbourhood because $H(K) \cap H(K')$ is the norm group of $KK' \subseteq L$. Separability: If $(a) \neq (b)$ then there exist $U((a))$ and $U((b))$ such that $U((a)) \cap U((b)) = 0$, because $(a)(b)^{-1} = \mathbf{N}/\mathbf{N}$!

[13] Cf. again no. 7.

First we observe that $\chi(\mathbf{a})$ is an element of the Galois group $\mathbf{G}$, for $\chi_i(\bar{a}_i) = g_i$ is an element of the Galois group G_i belonging to the field K_i of the approximating sequence. Therefore there remains to be shown only that

$$\chi_{i-1}(\bar{a}_{i-1}) = \chi_i(\bar{a}_i)\psi_i \qquad \text{for every } i,$$

and this is a consequence of the property iii) of the norm residue symbol. We see now that

$$\chi_i(\bar{a}_i) = \chi_i(a_i) \to \chi_{i-1}(a_i) = \chi_{i-1}(\bar{a}_{i-1}N_{K_i/k}A_{i-1})$$
$$= \chi_{i-1}(\bar{a}_{i-1}).$$

Here A_{i-1} denotes an element of the field K_{i-1} such that the element a_i which is an arbitrary representative of the class $\bar{a}_i$, becomes equal to $a_{i-1}N_{K_{i-1}/k}A_{i-1}$ where a_{i-1} is a representative of the class $\bar{a}_{i-1}$. The assumption that $\mathbf{a}$ is a fundamental sequence guarantees the existence of such an element A_{i-1}.

The symbol $\chi(\mathbf{a})$ furnishes an isomorphic mapping of $\mathbf{k}$ on $\mathbf{G}$.

We must show that every element $\mathbf{g} = (g_1, g_2, \cdots, g_i, \cdots)$ of the group $\mathbf{G}$ can be obtained. Again we can refer to the fundamental properties of the finite norm residue symbol. By property iv) there always exist elements a_i of the field k such that $\chi_i(a_i) = g_i$ for given g_i and by virtue of property ii) the elements a_i are determined up to norms $N_{K_i/k}A_i$ of elements A_i in K_i.

We now form the system $(\bar{a}_1, \bar{a}_2, \cdots, \bar{a}_i, \cdots)$ where the $\bar{a}_i$ are the uniquely determined residue classes of the elements (a_i) modulo $H(K_i)/\mathbf{N}$. Then that infinite row is an element $\mathbf{a}$ of the adjoint group $\mathbf{k}$. We must show that

$$\bar{a}_i\varphi_i = \bar{a}_{i-1}.$$

But
$$g_i = \chi_i(\bar{a}_i) \to g_{i-1} = \chi_{i-1}(\bar{a}_{i-1}) = g_i\psi_i$$

by construction

and
$$g_i = \chi_i(\bar{a}_i) \to \chi_{i-1}(a_i)$$

by property iii) of the symbol. Hence $\chi_{i-1}(\bar{a}_{i-1}) = \chi_{i-1}((a_i))$ or (a_i) is congruent to $\bar{a}_{i-1}$ modulo $H(K_{i-1})/\mathbf{N}$, and this is equivalent to $\bar{a}_{i-1} = \bar{a}_i\varphi_i$.

We are now able to state some corollaries concerning the relations of the subfields F of L and the closed subgroups $\mathbf{h}$ of $\mathbf{k}$.

The existence of the isomorphism χ between the adjoint group $\mathbf{k}$ and the galois group $\mathbf{G}$ allows us to translate the fundamental theorem of Galois theory, namely that there is a one to one correspondence between the subfields F of L and the closed subgroups $\mathbf{F}$ of $\mathbf{G}$, into a theorem about subgroups of $\mathbf{k}$.

COROLLARY 1. *The closed subgroups* $\mathbf{h}$ *of* $\mathbf{k}$ *correspond uniquely to the subfields* F *of* L *and the Galois group of* F *over* k *is isomorphic to the factor group* $\mathbf{k}/\mathbf{h}$.

W. Krull has shown that a closed subgroup $\mathbf{h}$ of $\mathbf{k}$ is uniquely determined by its components h_i in $k^*/H(K_i)$.[14] In this case the components h_i are determined

[14] Cf. no. 7.

by the norm groups $H(K_i \cap F)$ corresponding to the intersections of the associated field F and the fields K_i of an approximating sequence of L by the relations $h_i = H(K_i \cap F)/H(K_i)$.

The intersection and compositum of any two closed subgroups $\mathbf{h}_1$ and $\mathbf{h}_2$ are again closed subgroups of $\mathbf{k}$. The first is obvious, the latter needs a more detailed consideration which may be found in the above mentioned paper of M. Moriya.[15]

COROLLARY 2. *If F_1 and F_2 are two subfields of L and $\mathbf{h}_1$ and $\mathbf{h}_2$ are their corresponding subgroups in $\mathbf{k}$, then*

$$F_1F_2 \leftrightarrow \mathbf{h}_1 \cap \mathbf{h}_2 \qquad \textit{and} \qquad F_1 \cap F_2 \leftrightarrow \mathbf{h}_1\mathbf{h}_2.$$

If we introduce the group $\mathbf{K}$ of all continuous characters of $\mathbf{k}$ and denote by $\mathbf{H}$ the group of characters which map the subgroup $\mathbf{h}$ of $\mathbf{k}$ into unity, then we can state both corollaries in a more direct form:[16]

COROLLARY 1′. *The subgroups $\mathbf{H}$ of the group of all characters of $\mathbf{k}$ correspond uniquely to the subfields F of L and $\mathbf{H}$ is the group of all characters belonging to the associated group of F.*

COROLLARY 2′. *If F_1 and F_2 are two subfields of L and $\mathbf{H}_1$ and $\mathbf{H}_2$ are the character groups of their associated groups, then*

$$F_1F_2 \leftrightarrow \mathbf{H}_1\mathbf{H}_2 \qquad \textit{and} \qquad F_1 \cap F_2 \leftrightarrow \mathbf{H}_1 \cap \mathbf{H}_2.$$

The set of all finite subfields K of an infinite abelian field L over k has the property that for any two subfields K and K' their compositum KK' and their intersection $K \cap K'$ belongs to it. The equivalent fact about the set of all norm groups $H(K)$ is that $H(K) \cap H(K')$ and $H(K)H(K')$ belong to it.

Now let $\mathfrak{S}'$ be a set $\{H(K)\}$ of subgroups of k which has the property just stated. Let n denote the finite index $k^*/H(K)$. The least common multiple of all the integers n is a certain G-number N. Let n_i be an arbitrary approximation of N, so that $\lim n_i = N$.

The factor groups k^*/k^{*n_i} are all finite. We form the intersection of all groups $H(K)$ that lie between k^* and k^{*n_i} and call it H_i. The groups H_i form a subset $\mathfrak{S}$ of $\mathfrak{S}'$. The intersections $\mathbf{N}'$ and $\mathbf{N}$ of all groups of $\mathfrak{S}'$ and $\mathfrak{S}$ are then equal by construction. The class group $k^*/\mathbf{N}$ can be topologized with respect to the approximations $k^*/H(K)$ and k^*/H_i and again both topologies are seen to be equivalent. Let $\mathbf{k}$ be the closure of $k^*/\mathbf{N}$ with respect to the system $\mathfrak{S}$.

THEOREM II. *For any closed group $\mathbf{k}$ belonging to a defining system $\mathfrak{S}$ of subgroups in k there exists an infinite abelian field L over k whose adjoint group $\mathbf{k}(L)$ is equal to the given group $\mathbf{k}$.*

[15] Cf. MII.

[16] L. Pontrjagin, *The general topological theorem of duality for closed sets*, also, *The theory of topological commutative groups*, Ann. of math. II, ser. 35 (1934).

Proof. The existence theorem of finite abelian extensions over the field k asserts that there exist abelian fields K_i over k whose Galois groups are isomorphic to k^*/H_i. They are uniquely determined. Since the groups H_i form a descending sequence, $\cdots \supseteq H_{i-1} \supseteq H_i \supseteq \cdots$, the fields K_i form an ascending sequence of fields, $\cdots \subseteq K_{i-1} \subseteq K_i \subseteq \cdots$. The sequence $\{K_i\}$ is the approximating sequence of an infinite abelian extension L of k. The field L is uniquely determined. It is obvious that the adjoint group $\mathbf{k}(L)$ is equal to $\mathbf{k}$. (Because there we have $H_i = H(K_i)$.)

Each abelian extension M of k whose degree m is a divisor of the G-number N may be called a field of exponent N.[17] Again let $\lim n_i = N$. Then we put $H_i = k^{*n_i}$. To this special set of subgroups belongs, as we have seen, a uniquely determined abelian field F. Then the following corollary holds and its proof is quite obvious from all the proceeding observations.

Corollary. *Every abelian field M of exponent N is contained in the field F and the fields M correspond uniquely to the subgroups of the character group of* $\mathbf{k}(F)$.

Remark. The unramified infinite extensions W of k which are all fields of roots of units correspond to the different topologies that can be introduced in the infinite cyclic group.[18] This follows from the fact that all units of the field k are norms in the case of unramified finite extensions.[19] Therefore $\mathbf{N}$ is equal to the group of all units and $k^*/\mathbf{N}$ is isomorphic to the infinite cycle $\{p\}$ generated by the prime element p of k.

For instance if we take as neighbourhoods of unity in $\{p\}$ the groups $U_i = \{p^{q^i}\}$ for $i = 1, 2, \cdots$ we obtain the additive group of all integer q-adic numbers as Galois group of the field of all $(P^{q^i} - 1) - st$ roots of unity. Here the number P denotes the number of residue classes of the field k, modulo its prime ideal (p), and q an arbitrary prime. Or, if we take $U_i = \{p^{n_1 n_2 \cdots n_i}\}$ where n_i denotes the i^{th} integer as neighbourhoods, we have the field W of all roots of unity; its group is isomorphic to a Cantor group or monothetic group.

The Institute for Advanced Study.

[17] We say m divides N if m can be approximated by divisors of n_i.

[18] For the theory of the unramified and cyclic finite extensions cf. H. Hasse, *Über p-adische Schiefkörper und ihre Bedeutung für die Arithmetik hyperkomplexer Zahlsysteme*, Math. Annalen vol. 104 (1930).

[19] Cf. no. 18.

ANNALS OF MATHEMATICS
Vol. 38, No. 2, April, 1937

NOTE ON MATRIC ALGEBRAS

BY HERMANN WEYL

(Received December 3, 1936)

1. **Preliminaries.** In §1 and Appendix 7 of my paper, *Generalized Riemann Matrices and Factor Sets,*[1] I established the theory of matric algebras by operating with the matrices themselves and their vector space rather than with abstract elements, and getting along without the discussion of the radical and its influence upon the structure of the algebra. One can treat the $\times$-multiplication in the same style and thus complete the theory, as I propose to show briefly in this note, in which I make use of the same notations and nomenclature as in the cited chapter. At the bottom we have a (commutative) field k; the words "in k" should tacitly be supplied to all terms like "matrix," "algebra," "irreducible."

The set of all d-rowed matrices (transformations in a d-dimensional vector space) is denoted by $\mathfrak{M}_d$ (complete matric algebra). $E = E_d$ is the unit matrix. The linear closure of a given matric set $\mathfrak{A} = \{A\}$ consisting of all possible linear combinations of the elements A of $\mathfrak{A}$ with coefficients in k will be indicated by

$$[\mathfrak{A}] \quad \text{or} \quad [A]_{A \text{ in } \mathfrak{A}}.$$

The set of matrices

$$\begin{Vmatrix} A & 0 & \cdots & 0 \\ 0 & A & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & A \end{Vmatrix} \quad (u \text{ rows}) \text{ and } \quad \begin{Vmatrix} A_{11} & \cdots & A_{1u} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ A_{u1} & \cdots & A_{uu} \end{Vmatrix}$$

where A or the A_{ik} vary independently in $\mathfrak{A}$ is called $u \cdot \mathfrak{A}$ and $\mathfrak{A}_u$ respectively; they are algebras if $\mathfrak{A}$ is such. By transition to a new coördinate system a set $\mathfrak{A}$ of linear mappings A passes into what is called an equivalent set ($\sim \mathfrak{A}$). When δ variables x_ι undergo a linear substitution A and d variables y_k undergo a linear substitution B, then the δd products $x_\iota y_k$ undergo the substitution $A \times B$. If the x_ι and y_k are looked upon as components of vectors x, y in a δ- and d-dimensional vector space $\mathfrak{x}$ and $\mathfrak{y}$ respectively, then $z_{\iota k} = x_\iota y_k$ are the components of a vector $z = xy$ in the δd-dimensional product space $\mathfrak{x}\mathfrak{y}$. If A ranges over a set $\mathfrak{A}$ and B over $\mathfrak{B}$, we mean by $\mathfrak{A} \times \mathfrak{B}$ the set of all matrices

$$A \times B, \qquad A \text{ in } \mathfrak{A}, \qquad B \text{ in } \mathfrak{B}.$$

[1] Annals of Math. 37 (1936), pp. 709–745.

This is not an algebra even if $\mathfrak{A}$ and $\mathfrak{B}$ are such. The linear closure

$$[A \times B]_{A \text{ in } \mathfrak{A}, B \text{ in } \mathfrak{B}}$$

is to be denoted by $[\mathfrak{A} \times \mathfrak{B}]$ and will be called the *algebra product* of the algebras $\mathfrak{A}$ and $\mathfrak{B}$. We have

$$\mathfrak{A}_v = [\mathfrak{M}_v \times \mathfrak{A}],$$

hence

$$[\mathfrak{A}_v \times \mathfrak{B}] = [\mathfrak{A} \times \mathfrak{B}]_v .$$

LEMMA (1-A). *Let $R = \{C\}$ be an algebra of transformations in a d-dimensional vector space $\mathfrak{r}$, containing E_d. The product $\rho\mathfrak{r}$ of $\mathfrak{r}$ with a vector space ρ of dimensionality δ may be considered as the substratum of the transformations of $\mathfrak{M}_\delta \times R$. Then each of its invariant subspaces is of form $\rho\mathfrak{r}'$ where $\mathfrak{r}'$ is an invariant subspace of $\mathfrak{r}$. In particular, irreducibility of R entails the same for R_δ.*

Of this lemma I made use in the proof of the criterion, Theorem (1.3-D), l.c. Its demonstration is fairly obvious. Let $\epsilon_1, \cdots, \epsilon_\delta$ be a basis of ρ. Each vector z in $\rho\mathfrak{r}$ may be decomposed according to

$$z = \epsilon_1 z_1 + \cdots + \epsilon_\delta z_\delta \qquad (z_\iota \text{ vector in } \mathfrak{r}).$$

$E_{\iota\kappa}$ being the δ-rowed matrix which has a 1 at the crossing point of the ι^{th} row and the κ^{th} column and 0 elsewhere, application of the operation $E_{\iota\iota} \times E$ shows that our invariant subspace $\bar{\mathfrak{r}}$ of $\rho\mathfrak{r}$ contains the parts $\epsilon_\iota z_\iota$ of each z in $\bar{\mathfrak{r}}$. The operations like $E_{12} \times E$ prove that with $\epsilon_1 z_1$ (z_1 in $\mathfrak{r}$) also $\epsilon_2 z_1$ lies in $\bar{\mathfrak{r}}$. Consequently $\bar{\mathfrak{r}}$ is of the form $\rho\mathfrak{r}'$, and the operators $E \times C$ force the subspace $\mathfrak{r}'$ of $\mathfrak{r}$ to be invariant.

An abstract *division algebra* $\rho = \{\gamma\}$ of order δ is irreducibly represented by associating with γ the substitution

$$\gamma^*: \qquad \xi' = \gamma\xi \qquad (\xi \text{ varying in } \rho)$$

(regular representation). $\rho^* = \{\gamma^*\}$. The substitution

$$\xi' = \xi\gamma$$

may be denoted by γ_* and the set of all γ_* by ρ_*. A division algebra is *normal* provided the centrum consists of the multiples of the unit only. In the Appendix I proved the following statement concerning normal division algebras by way of a simple application of Burnside's criterion:

LEMMA (1-B). *Let ρ be a normal division algebra. The δ^2 substitutions*

$$\xi' = \alpha\xi\beta$$

one obtains by letting α and β run independently over a basis $\epsilon_1, \cdots, \epsilon_\delta$ of ρ yield a basis for the complete matric algebra $\mathfrak{M}_\delta$.

One might put this down in the following formula

$$[\alpha^*\beta_*]_{\alpha,\beta \text{ in } \rho} = \mathfrak{M}_\delta .$$

2. **The basic argument.** We are now going to consider the product $\rho\mathfrak{r}$ of a normal division algebra ρ of order δ and a d-dimensional vector space $\mathfrak{r}$ subject to the transformations C of a given irreducible matric algebra $R = \{C\}$. The space $\rho\mathfrak{r}$ is considered the substratum of the transformation set

$$\rho^* \times R = \{\gamma^* \times C\}_{\gamma \text{ in } \rho,\, C \text{ in } R}.$$

We then maintain that $\rho\mathfrak{r}$ splits into a number u of *irreducible* invariant subspaces in each of which $\gamma^* \times C$ induces the *same* transformation; or that we have an equivalence

$$\rho^* \times R \sim u \cdot \mathfrak{H}, \qquad \mathfrak{H} \text{ irreducible.} \tag{2.1}$$

This statement and its proof are the backbone of our whole discussion; the rest is mere juggling around and interpretation of the result. We proceed as follows.

The vectors x of $\rho\mathfrak{r}$ are expressed in terms of a basis $e_1, \cdots, e_d$ of $\mathfrak{r}$ as:

$$x = \xi_1 e_1 + \cdots + \xi_d e_d \qquad (\xi_i \text{ in } \rho). \tag{2.2}$$

Considering ρ as a "quasi-field" in which the coefficients ξ_i vary freely, we define

$$\gamma x = (\gamma\xi_1)e_1 + \cdots + (\gamma\xi_d)e_d, \qquad x\gamma = (\xi_1\gamma)e_1 + \cdots + (\xi_d\gamma)e_d.$$

An invariant subspace $\mathfrak{l}$ of $\rho\mathfrak{r}$ is certainly a subset of vectors x of form (2.2), closed with respect to addition and front multiplication ($x \to \gamma x$); for the latter operation is what we formerly denoted by $\gamma^* \times E$. Hence $\mathfrak{l}$ has a ρ-basis $l_1, \cdots, l_n$ in terms of which every x in $\mathfrak{l}$ is uniquely expressible as

$$x = \eta_1 l_1 + \cdots + \eta_n l_n \qquad (\eta_i \text{ in } \rho),$$

and the dimensionality δn of $\mathfrak{l}$ is a multiple of δ.

β being a given quantity in ρ, the space $\mathfrak{l}\beta$ containing all vectors $x\beta$ (x in $\mathfrak{l}$) is invariant with respect to $\gamma^* \times C$ as well as $\mathfrak{l}$, and $\gamma^* \times C$ induces therein the same transformation as in $\mathfrak{l}$. By making use of a basis $\epsilon_1 = 1, \cdots, \epsilon_\delta$ of ρ we apply the "typical argument" to the row of irreducible invariant subspaces

$$\mathfrak{l}_1 = \mathfrak{l}\epsilon_1 = \mathfrak{l},\ \mathfrak{l}_2 = \mathfrak{l}\epsilon_2, \cdots, \mathfrak{l}_\delta = \mathfrak{l}\epsilon_\delta$$

and thus succeed in picking out a number among them which by a proper arrangement may be denoted by $\mathfrak{l}_1, \cdots, \mathfrak{l}_u$ such that 1) $\mathfrak{l}_1, \cdots, \mathfrak{l}_u$ are linearly independent, and 2) each $\mathfrak{l}_\iota$ ($\iota = 1, \cdots, \delta$) is contained in the sum

$$\mathfrak{l}_1 + \cdots + \mathfrak{l}_u = (\mathfrak{l}).$$

The latter fact shows that $(\mathfrak{l})$ is also invariant with respect to back multiplications: $(\mathfrak{l})\epsilon_\iota$ is contained in $(\mathfrak{l})$ for $\iota = 1, \cdots, \delta$. Consequently $(\mathfrak{l})$ is invariant with respect to all transformations of the type

$$\alpha^*\beta_* \times C, \qquad \alpha \text{ and } \beta \text{ varying over a basis of } \rho,\ C \text{ in } R,$$

and thus, according to Lemma (1-B), with respect to $\mathfrak{M}_s \times R$. Lemma (1-A) then proves (I) to be the total space $\rho\mathfrak{r}$, and this remark finishes our demonstration, at the same time yielding the equation

$$d = nu:$$

u is a divisor of d.

3. **Exploitation.** Here we restate Theorem (1.3-B) of my former paper as

LEMMA (3-A). *An irreducible matric algebra R is $\sim r_v^*$ where r is a division algebra (and v a natural number).*

From (2.1) there follow the equations

$$[\rho^* \times R] \sim u[\mathfrak{H}], \qquad [\rho_v^* \times R] \sim u[\mathfrak{H}]_v .$$

According to Lemma (1-A), the algebra $[\mathfrak{H}]_v$ is irreducible as well as $[\mathfrak{H}]$; and in view of Lemma (3-A) we may put our result into the equivalence

$$(3.1) \qquad [\mathrm{P} \times R] \sim u\cdot\mathfrak{P}, \qquad \mathfrak{P} \text{ irreducible},$$

holding for any two irreducible matric algebras P and R the first of which is normal. Our result implies the abstract statement:

THEOREM (3-B). *The algebra product of two simple algebras one of which is normal, is a simple algebra again.*

From this we could infer our concrete proposition that $[\mathrm{P} \times R]$ is a certain multiple u of an irreducible $\mathfrak{P}$ by means of the general fact mentioned on page 714, l.c., that every representation of a simple algebra is a multiple of its irreducible representation. Our proof here, however, aimed directly at this concrete statement and yielded the further result that u is a divisor of the degree d of R.

Lemma (3-A) makes transition from division algebras r to simple algebras R so easy that it is perhaps convenient to specialize our result (2.1) to the case $R = r^*$ rather than to generalize it to (3.1). Hence let us write down the (special $\times$ special)-equation

$$(3.2) \qquad \rho^* \times r^* \sim u\cdot\mathfrak{H}$$

This leads back to the (general $\times$ general)-result (3.1) in the form

$$(3.3) \qquad [\rho_v^* \times r_w^*] \sim u\cdot[\mathfrak{H}]_{vw} .$$

Transition from ρ^* and r^* to $\mathrm{P} = \rho_v^*$ and $R = r_w^*$ leaves the multiplicity u unchanged while replacing $[\mathfrak{H}]$ by the likewise irreducible $[\mathfrak{H}]_{vw}$.

Concerning the (special $\times$ special)-case (3.2), I feel bound to make two additional remarks.

First remark. An invariant subspace of $\rho\mathfrak{r}$ has a basis $l_1, \cdots, l_n$ relative to the quasi-field of coefficients in ρ. However, we may exchange the rôles of ρ and r and look upon ρ as a vector space and on r as a quasi-field of multi-

plicators or coefficients. $\mathfrak{l}$ will then have an r-basis $\lambda_1, \cdots, \lambda_\nu$ in terms of which

$$y_1\lambda_1 + \cdots + y_\nu\lambda_\nu$$

describes $\mathfrak{l}$ with the coefficients y_ι ranging over r. The dimensionality of $\mathfrak{l}$ is

$$n\delta = \nu d, \qquad \text{therefore} \qquad d:\delta = n:\nu.$$

As $d = nu$, we obtain the further relation $\delta = \nu u$ and thus realize that *u is a common divisor of d and δ.* The same relationship prevails in the (general $\times$ general)-case (3.1). For in passing from ρ^* to $\mathrm{P} = \rho_v^*$ and from r^* to $R = r_w^*$, the degrees δ and d change into δv and dw respectively, while u stays put, eq. (3.3). With this additional information on hand we give the concrete counterpart of Theorem (3-B) as follows:

THEOREM (3-C). *The algebra product of two irreducible matric algebras P and R one of which is normal, decomposes into a number u of equal irreducible components $\mathfrak{P}$ according to the equivalence*

$$[\mathrm{P} \times R] \sim u \cdot \mathfrak{P}.$$

The multiplicity u is a common divisor of the degrees of both factors.

The u equal parts into which the generic matrix $\Gamma \times C$ of $\mathrm{P} \times R$ decomposes will occasionally be denoted by $\Pi(\Gamma, C)$. In the special case $\mathrm{P} = \rho^*$ we simply write $\Pi(\gamma, C)$ instead of $\Pi(\gamma^*, C)$, and similarly when R is specialized into r^*.

Second remark. In the (special $\times$ special)-relation (3.2) or in

$$[\rho^* \times r^*] \sim u[\mathfrak{H}]$$

we apply Lemma (3-A) to the irreducible $[\mathfrak{H}]$ and infer from it that

$$[\mathfrak{H}] = \mathfrak{p}_v^*$$

where the abstract division algebra $\mathfrak{p}$, called the R. Brauer product, is uniquely determined by the factors ρ and r. Comparison of degrees and orders in the ensuing equivalence

$$[\rho^* \times r^*] \sim u \cdot \mathfrak{p}_v^*$$

leads to the relations

$$\delta d = uv\mathfrak{d}, \qquad \delta d = v^2\mathfrak{d},$$

$\mathfrak{d}$ being the degree of $\mathfrak{p}^*$ = order of $\mathfrak{p}$. Hence $v = u$ and

$$d = nu, \qquad \delta = \nu u, \qquad \mathfrak{d} = n\nu:$$

THEOREM (3-D). *The algebra product of the regular representations ρ^* and r^* of two division algebras ρ and r of orders δ and d decomposes according to*

$$[\rho^* \times r^*] \sim u \cdot \mathfrak{p}_u^*$$

provided ρ is normal. Putting

$$d = nu, \qquad \delta = \nu u, \qquad (n \text{ and } \nu \text{ integers})$$

the order of the division algebra $\mathfrak{p}$ equals $n\nu$.

4. **Adjunction.** We have not as yet evaluated to the full the idea involved in our backbone proof that an invariant subspace $\mathfrak{l}$ of $\rho\mathfrak{r}$ can be referred to a ρ-basis $l_1, \cdots, l_n$. Let us now consider its implications for the case which is the other way around: $\mathrm{P} \times r^*$, P being a normal irreducible matric algebra, r an arbitrary division algebra. Let $\mathfrak{l}$ be an invariant subspace of the vector space $\mathfrak{r}r$ upon which the operators $\Gamma \times c^*$ of $\mathrm{P} \times r^*$ work (Γ are operators in the δ-dimensional space $\mathfrak{r}$, r is of order d). $\mathfrak{l}$ has an r-basis $\mathbf{l}_1, \cdots, \mathbf{l}_\nu$ such that each $\mathbf{x}$ of $\mathfrak{l}$ is uniquely expressible as

$$\mathbf{x} = y_1\mathbf{l}_1 + \cdots + y_\nu\mathbf{l}_\nu \qquad (y_\iota \text{ in } r). \tag{4.1}$$

We now look upon $\mathfrak{r}r = \mathfrak{r}_r$ as the vector space $\mathfrak{r}$ under extension of its field of multiplicators k into the quasi-field r. The elements of $\mathfrak{r}_r$ are rows $\mathbf{x}$ of δ quantities $x_1, \cdots, x_\delta$ in r. Addition is defined in the obvious manner, multiplication by a quantity c of r as:

$$c(x_1, \cdots, x_\delta) = (cx_1, \cdots, cx_\delta).$$

A linear subspace $\mathfrak{l}_r$ of $\mathfrak{r}_r$ is a subset closed with respect to addition and multiplication by any c in r. The subspace $\mathfrak{l}_r$ has a basis $\mathbf{l}_1, \cdots, \mathbf{l}_\nu$ as indicated by eq. (4.1).

Each Γ is a linear substitution with ordinary numbers $\gamma_{\iota\kappa}$ in k:

$$x'_\iota = \sum_\kappa x_\kappa \gamma_{\iota\kappa} \qquad (\iota, \kappa = 1, \cdots, \delta)$$

and hence commutes with all the multiplications $\mathbf{x} \to c\mathbf{x}$. If $\mathfrak{l}_r$ is invariant with respect to the transformations Γ of P, then each Γ: $\mathbf{x} \to \mathbf{x}'$ carries the basic vectors $\mathbf{l}_1, \cdots, \mathbf{l}_\nu$ into linear combinations of themselves:

$$\mathbf{l}'_\iota = \sum_\kappa c_{\iota\kappa}\mathbf{l}_\kappa \qquad (\iota, \kappa = 1, \cdots, \nu).$$

Commuting as it does with the multiplications, Γ then carries (4.1) into

$$\mathbf{x}' = y_1\mathbf{l}'_1 + \cdots + y_\nu\mathbf{l}'_\nu = y'_1\mathbf{l}_1 + \cdots + y'_\nu\mathbf{l}_\nu$$

where

$$y'_\iota = \sum_\kappa y_\kappa c_{\iota\kappa} \qquad (\iota, \kappa = 1, \cdots, \nu).$$

Here it is quite essential to write the coefficients $c_{\iota\kappa}$ *after* or to the right of the variables y_κ: we therefore speak of a *right-transformation.* Γ induces in $\mathfrak{l}_r$ the right-transformation $\| c_{\iota\kappa} \|$ and the correspondence $\Gamma \to \| c_{\iota\kappa} \|$ constitutes a *right-representation* of P in r. It seems worth while to present our chief result in this new garb:

THEOREM (4-A). *Under extension of k into a quasi-field r over k, a given normal matric algebra P, irreducible in k, breaks up into u equal irreducible right-representations of P in r.*

This mode of visualizing the situation is related to our former viewpoint in the following manner. Adopting the coördinate system here used, $\Pi(\Gamma, \mathfrak{l})$ arises from our right-representation $\Gamma \to \| c_{\iota\kappa} \|$ in replacing each $c_{\iota\kappa}$ by the

matrix $(c_{\iota\kappa})_*$ (back multiplication and hence lower asterisk!) whereas $\Pi(E_\delta, c)$ is simply $E_v \times c^*$.

Of particular import is the case when r is a commutative field K. We then deduce from our theorem that a normal irreducible matric algebra P in k splits into u equal irreducible matric algebras in K after extending the reference field k to a finite field K over k. Let us consider again the special case $\mathrm{P} = \rho^*$. We then must have an equivalence like

$$\rho^* \text{ ext. to } K \sim u \cdot \pi_v^*$$

where π is a (normal) division algebra in K. θ being the degree = order of π^*, comparison of degrees and orders leads to the relations

$$\delta = uv\theta, \qquad \delta = v^2\theta,$$

hence

$$u = v \quad \text{and} \quad \delta = u^2\theta.$$

THEOREM (4-B). *A normal division algebra ρ in k breaks up according to the equation*

$$\rho^* \text{ ext. to } K \sim u \cdot \pi_u^*$$

under extension of the field k into a finite field K over k. Hence

$$(\text{order of } \rho) = u^2 \cdot (\text{order of } \pi).$$

An easy consequence thereof is our final

THEOREM (4-C). *The order of a normal division algebra is a square number.*

Indeed, Theorem (4-B) informs us that on successive extensions of the reference field the order of ρ shrinks by throwing off square factors. Therefore we merely have to show how to find an extension so as to effect actual reduction ($u > 1$) as long as one has not yet reached the core: $\rho = (\mathrm{I})$. If ρ is not of order 1 we choose an element α of ρ different from any multiple of the unit and adjoin a root z of the characteristic equation $\varphi(z) = 0$ of the substitution $\alpha_*: \xi \to \xi\alpha$. The transformation $\alpha_* - zE$ is then singular, $\neq 0$, and commutes with all γ^*; hence, according to Schur's Lemma, ρ^* must needs reduce in $K = k(z)$. What one actually does is to pick out an irreducible k-factor $\psi(z)$ of $\varphi(z)$ and then define $k(z)$ *in abstracto* as the field of all k-polymonials of the indeterminate z modulo $\psi(z)$.

Here we find ourselves at the entrance gate to the "*splitting fields*," and there we might well finish our brief journey over what seems to me a particularly smooth and open road through this well-explored territory.

It was essential to suppose one of our factors to be normal. Again, the unruly things happen in the commutative fields: the algebra product of two fields breaks up into inequivalent parts, and to secure its full reducibility at all assumptions concerning separability are needed.

THE INSTITUTE FOR ADVANCED STUDY

ANNALS OF MATHEMATICS
Vol. 38, No. 2, April, 1937

PSEUDO-LINEAR TRANSFORMATIONS[1]

BY N. JACOBSON

(Received May 29, 1936)

It is well known that two matrices T_1 and T_2 with coördinates in a commutative field are similar, i.e. there exists a non-singular matrix A such that $T_2 = A^{-1}T_1A$ if and only if they have the same elementary divisors. In a number of problems occurring in complex projective geometry, differential and difference equations, and hypercomplex numbers a more general problem of similarity is encountered: namely, when are two matrices with coördinates in a non-commutative field $\mathfrak{F}$ related in the fashion

$$(*) \qquad T_2 = A^{-1}T_1A^S + A^{-1}A_S$$

where $A = (\alpha_{ij})$ is a matrix in $\mathfrak{F}$, $A^S = (\alpha_{ij}^S)$, $A_S = (\alpha_{ijS})$ and the α^S and α_S are obtained respectively from the α by a fixed 1-automorphism and differentiation in $\mathfrak{F}$? (§1 for definitions). In analogy with the special case of matrices in a commutative field under ordinary similarity, we define the elementary divisors of a matrix subject to the transformation (*) as certain non-commutative polynomials and prove that a necessary and sufficient (n.a.s.c.) that T_1 and T_2 be similar is that their elementary divisors be similar (§8).

As in the usual theory the present notion of similarity carries with it notions of reducibility, decomposability, and complete reducibility (§3). Criteria for these may be given in terms of the elementary divisors. In the important special case of finite matrices these criteria may be sharpened (§9).

We consider also the automorphism ring $\mathfrak{A}$ of T, i.e. the ring of all matrices A such that $AT = TA^S + A_S$. This is a generalization of the ring of matrices commutative with T and has many interesting special cases. For example, cyclic algebras and the invariant ring of a differential polynomial as defined by Ore are included here. The general theory of the ring $\mathfrak{A}$ may be deduced from results due to v. d. Waerden, Fitting and others. A more detailed discussion is given of some of the special cases which have application to hypercomplex numbers and differential equations.

In the body of the paper we take the abstract point of view, formulating the problems in terms of linear transformations or groups with operators. The concepts of similarity, reducibility, etc. appear naturally as special cases of the concepts of isomorphism, existence of allowable subgroups, etc. of groups with operators. This procedure has the added advantage of enabling us to apply directly the general theorems on groups to our case. The reader is

[1] An abstract of this paper appeared in the Proc. Nat. Acad. Sci. [7]. A portion of this work was done by the author as National Research Fellow at the University of Chicago.

referred to v. d. Waerden ([15] vol. I, pp. 132–144 and II, Chap. 15) for a discussion of these notions.

1. **Definition of pseudo-linear transformations.** Let $\mathfrak{R}$ be a vector space of finite dimensionality n over an arbitrary field $\mathfrak{F}$ (not necessarily commutative). A *pseudo-linear transformation* (p.l.t.) $\mathbf{T}$ of $\mathfrak{R}$ is defined by the following conditions:

$$(x + y)\mathbf{T} = x\mathbf{T} + y\mathbf{T} \qquad x, y \in \mathfrak{R} \tag{1}$$

$$(x\alpha)\mathbf{T} = (x\mathbf{T})\alpha^S + x\alpha_S \tag{2}$$

where the correspondence $\alpha \to \alpha^S$ is a 1-*automorphism* and $\alpha \to \alpha_S$ an *S-differentiation* in $\mathfrak{F}$, i.e.

$$(\alpha + \beta)^S = \alpha^S + \beta^S \qquad (\alpha\beta)^S = \alpha^S\beta^S \tag{3}$$

$$(\alpha + \beta)_S = \alpha_S + \beta_S \qquad (\alpha\beta)_S = \alpha_S\beta^S + \alpha\beta_S \tag{4}$$

and $\alpha \to \alpha^S$ is $(1 - 1)$. Equation (1) states that $\mathbf{T}$ is an automorphism of the abelian group $\mathfrak{R}$ and (2) gives the commutation relation

$$\alpha\mathbf{T} = \mathbf{T}\alpha^S + \alpha_S \tag{2'}$$

between $\mathbf{T}$ and the automorphism $x \to x\alpha$ of $\mathfrak{R}$. We note the following examples of p.l.t.

(a) $\alpha^S \equiv \alpha$, $\alpha_S \equiv 0$. In this case $\mathbf{T}$ is a linear transformation (l.t.) in a vector space over any (non-commutative) field. The case $\mathfrak{F}$ commutative is classical.

(b) $\mathfrak{F}$ the field of complex numbers, α^S the conjugate complex of α, $\alpha_S \equiv 0$. P.l.t. of this type were first defined by Segre[2] and are of interest in complex projective geometry. If we suppose that $\mathfrak{F}$ is any field, $\alpha \to \alpha^S$ any 1-automorphism, $\alpha_S \equiv 0$, we obtain as a generalization of the transformations of Segre a type of p.l.t. which we shall call *semi-linear* (s.l.t.). Thus the transformation $x \to x\mu \equiv x\mathbf{M}$ is an s.l.t. with automorphism $\alpha \to \alpha^S = \mu^{-1}\alpha\mu$.

(c) The differential equations

$$u_i' = \frac{du_i}{d\tau} = \sum_{j=1}^{n} u_j\alpha_{ji}(\tau) \qquad (i = 1, \cdots, n) \tag{5}$$

where the $\alpha(\tau)$ are analytic functions of τ, define a p.l.t. in the following way: For $\mathfrak{R}$ we take the space of forms $u = \sum u_i\alpha_i(\tau)$ in the independent vectors u_i with coördinates $\alpha_i(\tau)$ in the field $\mathfrak{F}$ of analytic functions of τ. We define $\mathbf{T}$ by

$$u \to u\mathbf{T} = \sum u_i'\alpha_i(\tau) + \sum u_i\alpha_i'(\tau)$$

where u_i is given by (5).[3] $\mathbf{T}$ is easily seen to be pseudo-linear with the identity as its automorphism. More generally if $\mathfrak{F}$ is any field $\alpha^S \equiv \alpha$ we shall call the corresponding p.l.t. *differential* (d.t.).

[2] C. Segre [14] and v. d. Waerden [16] p. 4.

[3] Cf. Krull [8] p. 188.

2. **The matrices of a p.l.t.** Let $(e_1, \cdots, e_n)$ be a basis for $\mathfrak{R}$ and suppose that the p.l.t. $\mathbf{T}$ sends e_k into $e_k\mathbf{T}$ where

$$e_k\mathbf{T} = e_1\tau_{1k} + e_2\tau_{2k} + \cdots + e_n\tau_{nk} \qquad (k = 1, \cdots, n)$$

or, using the usual rule of matrix multiplication

$$(e_1\mathbf{T}, \cdots, e_n\mathbf{T}) = (e_1, \cdots, e_n)T \qquad T = (\tau_{ij}). \tag{6}$$

If x is the vector $\sum e_i\xi_i \equiv \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{pmatrix}$, then according to (1) and (2)

$$x\mathbf{T} = T\begin{pmatrix} \xi_1^S \\ \xi_2^S \\ \vdots \\ \xi_n^S \end{pmatrix} + \begin{pmatrix} \xi_{1S} \\ \xi_{2S} \\ \vdots \\ \xi_{nS} \end{pmatrix} \tag{7}$$

Thus the p.l.t. $\mathbf{T}$ is completely determined by its matrix in a fixed coördinate system, its automorphism $\alpha \to \alpha^S$ and its differentiation $\alpha \to \alpha_S$. Conversely, if T is any matrix and $\alpha \to \alpha^S$ and $\alpha \to \alpha_S$ any 1-automorphism and differentiation, the correspondence $x \to x\mathbf{T}$ defined by (7) is a p.l.t. We restrict ourselves in the present paper to the study of p.l.t.'s with the same automorphism and differentiation and so, for the sake of simplicity, we write $\alpha^S \equiv \bar{\alpha}$, $\alpha_S \equiv \alpha'$.

Let $(e_1^*, \cdots, e_n^*)$ be a second basis for $\mathfrak{R}$ related to $(e_1, \cdots, e_n)$ by

$$(e_1^*, \cdots, e_n^*) = (e_1, \cdots, e_n)A \qquad (e_1, \cdots, e_n) = (e_1^*, \cdots, e_n^*)A^{-1}$$

where $A = (\alpha_{ij})$. If T and T^* are respectively the matrices of $\mathbf{T}$ relative to the bases $(e_1, \cdots, e_n)$ and $(e_1^*, \cdots, e_n^*)$ a simple computation shows that

$$T^* = A^{-1}T\bar{A} + A^{-1}A'. \tag{8}$$

3. **Similarity, reducibility, decomposability and complete reducibility.** If $\mathbf{T}$ is a p.l.t. we denote the ring of transformations[4] in $\mathfrak{R}$ generated by $\mathbf{T}$ and the elements of $\mathfrak{F}$ by $\mathfrak{F}[\mathbf{T}]$. From the definition (1) and (2) it follows that the elements of $\mathfrak{F}[\mathbf{T}]$ may be represented as polynomials in $\mathbf{T}$ with coefficients (on the right of the powers of $\mathbf{T}$) in $\mathfrak{F}$. We shall consider $\mathfrak{R}$ as a group $(\mathfrak{R}, \mathfrak{F}[\mathbf{T}])$ with operators $\mathfrak{F}[\mathbf{T}]$.

We say that the p.l.t.'s $\mathbf{T}_1$ and $\mathbf{T}_2$ are *similar* if $(\mathfrak{R}, \mathfrak{F}[\mathbf{T}_1])$ is isomorphic as a group with operators to $(\mathfrak{R}, \mathfrak{F}[\mathbf{T}_2])$ i.e., if there exists a $(1 - 1)$ mapping $x \to x\mathbf{A}$ of $\mathfrak{R}$ on itself such that

$$(x + y)\mathbf{A} = x\mathbf{A} + y\mathbf{A} \qquad (x\mathbf{O}_1)\mathbf{A} = (x\mathbf{A})\mathbf{O}_2 \tag{9}$$

[4] The sum $\mathbf{T}_1 + \mathbf{T}_2$ of two transformations is defined as $x(\mathbf{T}_1 + \mathbf{T}_2) \equiv x\mathbf{T}_1 + x\mathbf{T}_2$, the product as $x(\mathbf{T}_1\mathbf{T}_2) \equiv (x\mathbf{T}_1)\mathbf{T}_2$.

where $\mathbf{O}_k = \mathbf{T}_k^m \alpha_0 + \mathbf{T}_k^{m-1}\alpha_1 + \cdots + \alpha_m$ ($k = 1, 2$). In order that (9b) hold it is sufficient with (9a) to require merely that $(x\alpha)\mathbf{A} = (x\mathbf{A})\alpha$ and $(x\mathbf{T}_1)\mathbf{A} = (x\mathbf{A})\mathbf{T}_2$. Thus $\mathbf{A}$ is a non-singular l.t. such that

$$\mathbf{T}_1\mathbf{A} = \mathbf{A}\mathbf{T}_2, \qquad \mathbf{T}_2 = \mathbf{A}^{-1}\mathbf{T}_1\mathbf{A}. \tag{10}$$

It follows readily that the matrices T_1, T_2, A of $\mathbf{T}_1$, $\mathbf{T}_2$, $\mathbf{A}$ satisfy the equation

$$T_1 = A^{-1}T_2\bar{A} + A^{-1}A'. \tag{11}$$

Conversely, if the matrices of two p.l.t. (with the same automorphism and differentiation, are related according to (11), the p.l.t. are similar.

If $\mathfrak{R}_1$ is an allowable subgroup, or *invariant subspace* of $\mathfrak{R}$, i.e. a subspace which is transformed into itself by $\mathbf{T}$, then $\mathbf{T}$ induces a p.l.t. $\mathbf{T}_1$ in $\mathfrak{R}_1$ which may then also be considered as a group with operators $(\mathfrak{R}_1, \mathfrak{F}[\mathbf{T}_1])$. We call $\mathbf{T}_1$ a *contraction* and denote its relation to $\mathbf{T}$ by $\mathbf{T} \geq \mathbf{T}_1$, or $\mathbf{T}_1 \leq \mathbf{T}$. In this case $\mathbf{T}$ also induces a p.l.t. in the *projection space* (factor group) $\mathfrak{R}/\mathfrak{R}_1$ which we denote as $\mathbf{T}/\mathbf{T}_1$. If $\mathfrak{R}_1$ and $\mathfrak{R}_2$ are two invariant subspaces, then so is their vector sum $\mathfrak{R}_1 \cup \mathfrak{R}_2$ and their intersection $\mathfrak{R}_1 \cap \mathfrak{R}_2$. If $\mathbf{T}_1$ and $\mathbf{T}_2$ are the contractions of $\mathbf{T}$ in $\mathfrak{R}_1$ and $\mathfrak{R}_2$, we denote the p.l.t.'s corresponding to $\mathfrak{R}_1 \cup \mathfrak{R}_2$ and $\mathfrak{R}_1 \cap \mathfrak{R}_2$ by $\mathbf{T}_1 \cup \mathbf{T}_2$ and $\mathbf{T}_1 \cap \mathbf{T}_2$ respectively. The set of contractions of $\mathbf{T}$ considered relative to the operations $\cup$ and $\cap$ is an instance of an abstract system of the type called a modular lattice by G. Birkhoff and a Dedekind structure by O. Ore.[5] Its characteristic property is the Dedekind law:

If $\mathbf{T}_1 \geqq \mathbf{T}_3$, then $\mathbf{T}_1 \cap (\mathbf{T}_2 \cup \mathbf{T}_3) = (\mathbf{T}_1 \cap \mathbf{T}_2) \cup \mathbf{T}_3$.

If $\mathbf{T} > \mathbf{T}_1$ ($\mathbf{T} \geqq \mathbf{T}_1$ but $\mathbf{T} \neq \mathbf{T}_1$) where $\mathbf{T}_1 \neq 0$, then $\mathbf{T}$ is *reducible*. If a basis of $\mathfrak{R}_1$ is supplemented to give a basis for $\mathfrak{R}$, it follows that the matrix of $\mathbf{T}$ relative to this basis has the form

$$\begin{pmatrix} T_1 & Q \\ O & T_2 \end{pmatrix} \tag{12}$$

where T_1 is a matrix of $\mathbf{T}_1$ and T_2 of $\mathbf{T}/\mathbf{T}_1$. The sequence of contractions

$$\mathbf{T} = \mathbf{T}_1 > \mathbf{T}_2 > \cdots > \mathbf{T}_l > \mathbf{T}_{l+1} = 0$$

is a *composition series* for $\mathbf{T}$ if each $\mathbf{T}_i/\mathbf{T}_{i+1}$ is irreducible. l is the *length* of the series and $\mathbf{T}_i/\mathbf{T}_{i+1}$ ($i = 1, \cdots, l$) the *composition factors*. The Jordan-Hölder theorem[6] asserts that any two composition series for $\mathbf{T}$ have the same length and that their composition factors are similar in pairs. $\mathbf{T}$ is *decomposable*, $\mathbf{T} = \mathbf{T}_1 \oplus \mathbf{T}_2$, if $\mathbf{T} = \mathbf{T}_1 \cup \mathbf{T}_2$ where $\mathbf{T}_1 \cap \mathbf{T}_2 = 0$ and $\mathbf{T}_i \neq 0$. If a basis for $\mathfrak{R}$ is chosen to consist of a basis for $\mathfrak{R}_1$ plus a basis for $\mathfrak{R}_2$, then $\mathbf{T}$ has a matrix of the form (12) with $Q = 0$. Conversely, if $\mathbf{T}$ has a matrix of the form (12) ((12) with $Q = 0$), then $\mathbf{T}$ is reducible (decomposable). The decomposition of $\mathbf{T}$ may be continued to obtain $\mathbf{T} = \mathbf{T}_1 \oplus \mathbf{T}_2 \oplus \cdots \oplus \mathbf{T}_u$ where

[5] G. Birkhoff [2] p. 445 and Ore [13] p. 411.

[6] v. d. Waerden [15] I, p. 140.

each $\mathbf{T}_i$ is indecomposable. In this connection we have the theorem of Krull[7] that any two decompositions of $\mathbf{T}$ into indecomposable parts have the same number of parts which may be paired off into similar pairs. If the indecomposable parts of $\mathbf{T}$ are irreducible, then $\mathbf{T}$ is said to be *completely reducible.* It follows that $\mathbf{T}$ has a matrix of the form

$$\begin{pmatrix} T_1 & & & \\ & T_2 & & \\ & & \ddots & \\ & & & T_u \end{pmatrix}$$

where the T_i are irreducible and conversely.

4. **Non-commutative polynomials.** Let $\mathfrak{F}[t, \bar{\ }, '] \equiv \mathfrak{F}[t]$ be the ring of non-commutative polynomials in the indeterminate t with coefficients taken on the right in $\mathfrak{F}$ and in which multiplication is defined by

$$\alpha t = t\bar{\alpha} + \alpha' \tag{13}$$

$\mathfrak{F}[t]$ is a domain of integrity with Euclidean algorithm. The ring $\mathfrak{F}[\mathbf{T}]$ associated with the p.l.t. is in a sense a representation of $\mathfrak{F}[t]$ and so it is natural to expect that the properties of the latter will be useful in studying $\mathbf{T}$. In this section we shall give a resumé of those properties of $\mathfrak{F}[t]$ which will be required below referring the reader to a paper of Ore's for a complete discussion.[8]

From the Euclidean algorithm follows the existence of a h.c.l.f. (h.c.r.f. $(\alpha(t), \beta(t))_L$ $((\alpha(t), \beta(t))_R)$ and a l.c.r.m. (l.c.l.m.) $[\alpha(t), \beta(t)]_R$ $([\alpha(t), \beta(t)]_L)$ of any two polynomials $\alpha(t)$, $\beta(t)$ in $\mathfrak{F}[t]$. We suppose these to be *normalized,* i.e. to have leading coefficient $= 1$. It follows that they are unique. The *left-transform* of $\alpha(t)$ by $\beta(t)$ is defined as

$$\alpha_1(t) \equiv \beta(t)^{-1}\cdot\alpha(t)\cdot\beta(t) = \beta(t)^{-1}[\alpha(t), \beta(t)]_R\bar{\beta}_0^{(r-d)}\alpha_0$$

where r is the degree $\sigma\alpha(t)$, $d = \sigma(\alpha(t), \beta(t))_R$, α_0 and β_0 the leading coefficients of $\alpha(t)$ and $\beta(t)$ respectively. Since

$$\sigma\alpha(t) + \sigma\beta(t) = \sigma[\alpha(t), \beta(t)]_R + \sigma(\alpha(t), \beta(t))_L$$

we have $\sigma\alpha_1(t) = r - d$ and so the factor $\bar{\beta}_0^{(r-d)}\alpha_0$ gives $\alpha_1(t)$ the leading coefficient α_0. If $d = 0$, $\sigma\alpha(t) = \sigma\alpha_1(t)$ and then $\alpha(t)$ and $\alpha_1(t)$ are said to be *left-similar.* We note the following rules for transforms:

$$\beta_1(t)^{-1}\cdot(\beta_2(t)^{-1}\cdot\alpha(t)\cdot\beta_2(t))\cdot\beta_1(t) = (\beta_2(t)\beta_1(t))^{-1}\cdot\alpha(t)\cdot(\beta_2(t)\beta_1(t)) \tag{14}$$

$$\beta(t)^{-1}\cdot[\alpha_1(t), \alpha_2(t)]_R\cdot\beta(t) = [\beta(t)^{-1}\cdot\alpha_1(t)\cdot\beta(t), \beta(t)^{-1}\cdot\alpha_2(t)\cdot\beta(t)]_R \tag{15}$$

$$\beta(t)^{-1}\cdot\alpha_1(t)\alpha_2(t)\cdot\beta(t) \equiv 0\ (\beta(t)^{-1}\cdot\alpha_1(t)\cdot\beta(t))_R. \tag{16}$$

[7] Krull [8] p. 186.

[8] Ore [12].

Similarly we define right-transforms and right-similarity. Right-similarity and left-similarity are coexistential and so we shall speak simply of *similarity* without the modifiers and denote it by the symbol $\simeq$. If $\mathfrak{F}[t]$ is commutative, i.e. $\mathfrak{F}$ is commutative and $\bar{\alpha} \equiv \alpha$, $\alpha' \equiv 0$, then similarity implies identity.

$\alpha(t)$ is *reducible* if it has a proper factor. If $\alpha(t) = [\alpha_1(t), \alpha_2(t)]_R$, $(\alpha_1(t), \alpha_2(t))_L = 1$ and $\sigma\alpha_i(t) > 0$, then $\alpha(t)$ is said to be *decomposable* into the *left-components* $\alpha_1(t)$, $\alpha_2(t)$. In this case $\sigma\alpha(t) = \sigma\alpha_1(t) + \sigma\alpha_2(t)$. If $\alpha(t) = \alpha_1(t)\beta_2(t) = \alpha_2(t)\beta_1(t)$, then $\alpha(t) = [\beta_1(t), \beta_2(t)]_L$, $(\beta_1(t), \beta_2(t))_R = 1$ and so $\alpha(t)$ is also decomposable into the *right-components* $\beta_1(t)$, $\beta_2(t)$ and conversely. $\alpha(t)$ is *completely reducible* if $\alpha(t) = [\pi_1(t), \cdots, \pi_u(t)]_R$ where the $\pi_i(t)$ are irreducible. It may be shown that this is equivalent to $\alpha(t) = [\rho_1(t), \cdots, \rho_u(t)]_L$ where the $\rho_i(t)$ are irreducible and that every factor of a completely reducible polynomials is also completely reducible.[9]

If the right-ideal $(\mu^*(t))_R$ (the right multiples of $\mu^*(t)$) is two sided, i.e. for every $\alpha(t)$ there exists a $\beta(t)$ such that

$$\alpha(t)\mu^*(t) = \mu^*(t)\beta(t) \tag{17}$$

then we say that $\mu^*(t)$ is *finite*. Another formulation of this condition is

$$\mu^*(t) \equiv 0 \ (\alpha(t)^{-1} \cdot \mu^*(t) \cdot \alpha(t))_R \tag{17'}$$

for every $\alpha(t)$. The correspondence $\mathbf{A}_{\mu^*}$: $\alpha(t) \to \beta(t)$ is a homomorphic mapping of $\mathfrak{F}[t]$ onto a subring $\mathfrak{G}[t_1]$ where $\mathfrak{F} \to \mathfrak{G}$, $t \to t_1$. However, it is easily seen that $\mathfrak{F} = \mathfrak{G}$. For the correspondent of α is $\beta = \rho^{-1}\bar{\alpha}^{(s)}\rho$ where $s = \sigma\mu^*(t)$, and ρ is the leading coefficient of $\mu^*(t)$ and so as α varies over the whole of $\mathfrak{F}$, so does β. Since t_1 is of the first degree in t, $\mathfrak{F}[t_1] = \mathfrak{F}[t]$ and so $\mathbf{A}_{r^*}$ is a 1-automorphism of $\mathfrak{F}[t]$. Thus for every $\beta(t)$ there corresponds an $\alpha(t)$ satisfying (17), or

$$\mu^*(t) \equiv 0 \ (\beta(t) \cdot \mu^*(t) \cdot \beta(t)^{-1})_L \tag{18}$$

for every $\beta(t)$.

If $\mathfrak{J}$ is a two-sided ideal in $\mathfrak{F}[t]$ and $\mu^*(t)$ an element of $\mathfrak{J}$ having minimum degree and leading coefficient 1, it is easily seen from the division process that every element of $\mathfrak{J}$ is both a r.m. and a l.m. of $\mu^*(t)$. Hence $\mu^*(t)$ is uniquely determined by $\mathfrak{J}$ and if $\alpha(t)$ is arbitrary $\alpha(t)\mu^*(t) = \mu^*(t)\beta(t)$, i.e. $\mu^*(t)$ is a normalized finite polynomial. Thus we have a $(1 - 1)$ correspondence between the two-sided ideals and the elements $\mu^*(t)$ and the theory of the latter which we shall develop here is equivalent to that of two-sided ideals.

If $\mu_1^*(t)$ and $\mu_2^*(t)$ are finite, then so is $\mu_3^*(t) = \mu_1^*(t)\mu_2^*(t)$ and $\mathbf{A}_{\mu_3^*} = \mathbf{A}_{\mu_1^*}\mathbf{A}_{\mu_2^*}$. If $\mu_3^*(t) = \mu_1^*(t)\mu_2^*(t)$ where $\mu_3^*(t)$ and $\mu_1^*(t)$ are finite, then so is $\mu_2^*(t)$. Let $\mathfrak{B}^*$ denote the totality of normalized finite elements. $\mathfrak{B}^*$ is closed under multiplication.

---

[9] Up to this point we have given an abstract of Ore's results. The rest of this section I believe is new.

If $\mu(t)$ is a l.f. of a $\mu^*(t) \in \mathfrak{B}^*$, say $\mu^*(t) = \mu(t)\nu(t)$, then $\mu(t)$ is *bounded.* Suppose

$$\delta(t) = (\mu(t), \mu^*(t))_R = \alpha_1(t)\mu(t) + \alpha_2(t)\mu^*(t),$$

then

$$\delta(t)\nu(t) = \mu^*(t)[\beta_1(t) + \beta_2(t)\nu(t)]$$

where $\beta_1(t)$, $\beta_2(t)$ are given by (17). Thus $\sigma\delta(t) = \sigma\mu(t)$ and $\mu^*(t) = \nu_1(t)\mu(t)$: *If $\mu(t)$ is a l.f. (r.f.) of a finite polynomial $\mu^*(t)$, then it is also a r.f. (l.f.) of $\mu^*(t)$.*

Let $\mu^*(t)$ be a polynomial of least degree in $\mathfrak{B}^*$ divisible by the bounded polynomial $\mu(t)$. Equation (16) shows that $\mu^*(t)$ is divisible by

$$\mu_1(t) = [\mu(t), \alpha(t)^{-1}\cdot\mu(t)\cdot\alpha(t), \beta(t)^{-1}\cdot\mu(t)\cdot\beta(t), \cdots]_R$$

where $1, \alpha(t), \beta(t), \cdots$ exhaust all the elements of $\mathfrak{F}[t]$. It follows from (14) and (15) that

$$\mu_1(t) \equiv 0 \ (\alpha(t)^{-1}\cdot\mu_1(t)\cdot\alpha(t))_R$$

for every $\alpha(t)$ and hence $\mu_1(t) \in \mathfrak{B}^*$. Since $\mu^*(t)$ was assumed to be minimal, $\mu_1(t) = \mu^*(t)$ and this polynomial is unique. We shall call it the *bound* of $\mu(t)$. It may be characterized also as

$$[\mu(t), \alpha(t)\cdot\mu(t)\cdot\alpha(t)^{-1}, \beta(t)\cdot\mu(t)\cdot\beta(t)^{-1}, \cdots]_L.$$

It is easily seen that the product of two bounded polynomials $\mu_1(t)\mu_2(t)$ is bounded with bound of degree $\sigma \leq \sigma\mu_1^*(t)\mu_2^*(t)$ where $\mu_i^*(t)$ is the bound of $\mu_i(t)$. We denote the multiplicative system of normalized bounded polynomials by $\mathfrak{B}$.

Lemma 1. *If $\pi(t)$ is an irreducible bounded polynomial, then its bound $\pi^*(t)$ is irreducible in $\mathfrak{B}^*$.*

If $\pi^*(t) = \pi_1^*(t)\pi_2^*(t)$, $\sigma\pi_i^*(t) > 0$, then $\pi(t)$ can not be a factor of $\pi_1^*(t)$ because of the minimality of $\pi^*(t)$. Hence $(\pi(t), \pi_1^*(t)) = 1$ and $\pi(t) = \pi_1^*(t)^{-1}\cdot\pi(t)\pi_1^*(t)$ is a factor of $\pi_2^*(t)$ again contrary to the minimality of $\pi^*(t)$.

Lemma 2. *$\mathfrak{B}^*$ is a commutative multiplicative system.*

Any polynomial in $\mathfrak{B}^*$ may be factored into a product of polynomials which are irreducible in $\mathfrak{B}^*$. It is therefore sufficient to prove that every pair of irreducible polynomials in $\mathfrak{B}^*$ are commutative. If $\pi_1^*(t) = \pi_2^*(t)$, this is evident. If $\pi_1^*(t) \neq \pi_2^*(t)$, then $(\pi_1^*(t), \pi_2^*(t)) = 1$ for otherwise their common factor would have two distinct bounds. Hence

$$\pi_1^*(t)\pi_2^*(t) = \pi_2^*(t)\rho_1^*(t) = [\pi_1^*(t), \pi_2^*(t)]_R$$

and similarly $\pi_2^*(t)\pi_1^*(t) = [\pi_1^*(t), \pi_2^*(t)]_R$. Thus $\pi_1^*(t)\pi_2^*(t) = \pi_2^*(t)\pi_1^*(t)$.

Corollary. *If $\mu_1^*(t)$ and $\mu_2^*(t) \in \mathfrak{B}^*$, then $\mathbf{A}_{\mu_1^*}\mathbf{A}_{\mu_2^*} = \mathbf{A}_{\mu_2^*}\mathbf{A}_{\mu_1^*}$.*

Theorem 1. *The polynomials in $\mathfrak{B}^*$ may be factored uniquely into irreducible polynomials in $\mathfrak{B}^*$.*

Let

$$\mu^*(t) = \pi_1^*(t)\pi_2^*(t) \cdots \pi_k^*(t) = \rho_1^*(t)\rho_2^*(t) \cdots \rho_l^*(t)$$

be two factorizations of $\mu^*(t)$ into factors irreducible in $\mathfrak{B}^*$. If $\pi(t)$ is a divisor of $\mu^*(t)$ irreducible in $\mathfrak{F}[t]$, then by the argument in Lemma 1, $\pi(t)$ is a divisor of one of the $\pi_i^*(t)$ and of one of the $\rho_1^*(t)$. Since we can permute the $\pi^*(t)$ and the $\rho^*(t)$, we may assume that $i = j = 1$. By the uniqueness of the bound of $\pi(t)$, we can conclude that $\pi_1^*(t) = \rho_1^*(t)$. We cancel off this factor and repeat the process to obtain the theorem.

We conclude this section with an explicit determination of the elements of $\mathfrak{B}^*$ in certain polynomial rings.

1. $\mathfrak{F}[t]$ *semi-linear*, i.e. has a generator t such that $\alpha t = t\bar{\alpha}$. Let

$$\mu^*(t) = t^m + t^{m-1}\mu_1 + \cdots + \mu_m \tag{19}$$

$\epsilon\, \mathfrak{B}^*$. The condition $t\mu^*(t) = \mu^*(t)t_1$ gives $t_1 = t$ and hence $\bar{\mu}_i = \mu_i$; $\alpha\mu^*(t) = \mu^*(t)\alpha_1$ implies $\alpha_1 = \bar{\alpha}^{(m)}$ and hence $\bar{\alpha}^{(m-i)}\mu_i = \mu_i\bar{\alpha}^{(m)}$. Thus if $\mu_i \neq 0$

$$\bar{\alpha}^{(m)} = \mu_i^{-1}\bar{\alpha}^{(m-i)}\mu_i \qquad \text{or} \qquad \bar{\alpha}^{(i)} = \mu_i^{-1}\alpha\mu_i$$

i.e. $\alpha \to \bar{\alpha}^{(i)}$ is an inner automorphism. If q is the smallest integer such that $\alpha \to \bar{\alpha}^{(q)}$ is an inner automorphism, then $i \equiv 0 \pmod{q}$. q will be called the *relative order* of the automorphism. It is easily seen that $\mu^*(t)$ has the form

$$\mu^*(t) = t^m + t^{m-q}\nu_1 + t^{m-2q}\nu_2 + \cdots \tag{20}$$

where

$$\bar{\nu}_j = \nu_j \qquad \nu_j\bar{\alpha}^{(jq)} = \alpha\nu_j \tag{21}$$

for all α. Conversely these conditions are sufficient to insure that $\mu^*(t) \,\epsilon\, \mathfrak{B}^*$.

THEOREM 2. *Let $\mathfrak{F}[t]$ be semi-linear. If the automorphism of $\mathfrak{F}[t]$ has relative infinite order, $\mathfrak{B}^*$ consists of the powers of t only. If the automorphism has relative finite order q, then* (20) *with* (21) *gives the form of the elements of $\mathfrak{B}^*$.*

2. $\mathfrak{F}[t]$ *differential*, i.e. $\alpha t = t\alpha + \alpha'$. If there exists a ρ such that $\alpha' = \rho\alpha - \alpha\rho$ for all α, then we shall call $\alpha \to \alpha'$ an *inner differentiation*. In this case we may replace t by $t_1 = t + \rho$ and obtain $\alpha t_1 = t_1\alpha$. $\mathfrak{F}[t]$ is then of the type considered in 1. We therefore assume that is an *outer differentiation*.

Let $\mu^*(t)$ of the form (19) be an element of $\mathfrak{B}^*$. $t\mu^*(t) = \mu^*(t)t_1$ implies $t_1 = t$ and $\mu_i' = 0$; $\alpha\mu^*(t) = \mu^*(t)\alpha_1$ implies $\alpha_1 = \alpha$ and since

$$\alpha t^j = t^j\alpha + \binom{j}{1}t^{j-1}\alpha' + \binom{j}{2}t^{j-2}\alpha'' + \cdots + \alpha^{(j)} \tag{22}$$

we have among other conditions

$$m\alpha' = \mu_1\alpha - \alpha\mu_1.$$

If the characteristic of $\mathfrak{F}$, $\chi(\mathfrak{F}) = 0$, it follows from the assumption that $\alpha \to \alpha'$ is outer that $m = 0$.

THEOREM 3. *If $\mathfrak{F}[t]$ is differential of characteristic* 0 *and with an outer differentiation, then $\mathfrak{B}^*$ contains the identity element only.*

If $\chi(\mathfrak{F}) = p \neq 0$, the conditions that $\mu^*(t) \in \mathfrak{B}^*$ are somewhat complicated but may be satisfied non-trivially. We hope to discuss this case in a later paper.

5. **Matrices with elements in** $\mathfrak{F}[t]$. Let $\mathfrak{F}[t]_n$[10] denote the ring of matrices of n rows and columns with coordinates in $\mathfrak{F}[t]$. If $U(t) \in \mathfrak{F}[t_n]$ and has an inverse $U^{-1}(t)$ in $\mathfrak{F}[t]_n$, then $U(t)$ is a *unit* in $\mathfrak{F}[t]_n$. If $B(t) = U(t)A(t)V(t)$ where $U(t)$, $V(t) \in \mathfrak{U}_n$ the multiplicative group of units, then $B(t)$ and $A(t)$ are *associates*. This relation is evidently reflexive, symmetric and transitive.

We consider the following problem: Given $A(t)$, to select an associate of $A(t)$ which has a simple normal form and to determine to what extent this normal form is unique. The result[11] given by Wedderburn and by van der Waerden in this connection is the following

LEMMA 3. *There exists an associate $B(t)$ of $A(t)$ having the form*

$$B(t) = \begin{pmatrix} \beta_1(t) & & & & & & \\ & \beta_2(t) & & & & & \\ & & \ddots & & & & \\ & & & \beta_r(t) & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{pmatrix} \tag{23}$$

where $\beta_i(t)$ is both a l.f. and a r.f. of $\beta_j(t)$ for $j > i$.

We shall prove an extension of this, namely

THEOREM 4. *The associate $B(t)$ of the diagonal form* (23) *may be chosen so that each $\beta_i(t)$ for $i < r$ is bounded and the bound $\beta_i^*(t)$ is a factor of $\beta_j(t)$ for $j > i$.*

(1) We begin with the associate $B(t)$ in (23). If every left-transform of $\beta_1(t)$ is a l.f. of $\beta_2(t)$, $\beta_1(t)$ is bounded and its bound $\beta_1^*(t)$ divides $\beta_2(t)$ and hence also $\beta_j(t)$, $j > 2$. Consider the matrix

$$B_{n-1}(t) = \begin{pmatrix} \beta_2(t) & & & & & \\ & \ddots & & & & \\ & & \beta_r(t) & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$$

[10] In general if $\mathfrak{A}$ is a ring $\mathfrak{A}_n$ will denote the ring of all matrices of n rows and columns with coördinates in $\mathfrak{A}$.

[11] Wedderburn [17] pp. 139–141 and v. d. Waerden [15] II pp. 120–125. Wedderburn's result is in fact an extension of Lemma 3. However, it too is included in Theorem 4.

of $\mathfrak{F}[t]_{n-1}$. Using induction on the number of rows we may conclude that $B_{n-1}(t)$ has an associate

$$C_{n-1}(t) = U_{n-1}(t)B_{n-1}(t)V_{n-1}(t) = \begin{pmatrix} \gamma_2(t) & & & & & \\ & \ddots & & & & \\ & & \gamma_r(t) & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}^{12}$$

where $\gamma_i^*(t)$ is a factor of $\gamma_j(t)$ for $i < j, < r$. Then

$$(24) \quad C(t) = \begin{pmatrix} 1 & \\ & U_{n-1}(t) \end{pmatrix} B(t) \begin{pmatrix} 1 & \\ & V_{n-1}(t) \end{pmatrix} = \begin{pmatrix} \gamma_1(t) & & & & & & \\ & \gamma_2(t) & & & & & \\ & & \ddots & & & & \\ & & & \gamma_r(t) & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{pmatrix}$$

where $\gamma_1(t) = \beta_1(t)$. Since $\gamma_i(t)$ for $i > 2$ are linear combinations of

$$\beta_2(t), \cdots, \beta_r(t),$$

it follows that $\gamma_1^*(t)$ is a factor of $\gamma_i(t)$ and hence the associate (24) of $A(t)$ has the required normal form.

(2) Suppose next that the transform $\lambda(t)^{-1}\cdot\beta_1(t)\cdot\lambda(t)$ is not a l.f. of $\beta_2(t)$. Then $\beta_1(t)$ is not a l.f. of $\lambda(t)\beta_2(t) = \xi(t)$. Let $\delta(t) = (\beta_1(t), \xi(t))_L$ and $\beta_1(t) = \delta(t)\mu(t)$, $\xi(t) = \delta(t)\nu(t)$ where $(\mu(t), \nu(t))_L = 1$. It follows that $\eta_1(t)$, $\zeta_1(t)$, $\eta_2(t)$, $\zeta_2(t)$ can be determined so that

$$\mu(t)\eta_1(t) + \nu(t)\zeta(t) = 1 \qquad \mu(t)\eta_2(t) + \nu(t)\zeta_2(t) = 0$$

and so that

$$W_2(t) = \begin{pmatrix} \eta_1(t) & \eta_2(t) \\ \zeta_1(t) & \zeta_2(t) \end{pmatrix}$$

---

[12] The rank $r - 1$ is invariant: Wedderburn [17] p. 140.

is a unit in $\mathfrak{F}[t]_2$.[13] Thus

$$
\begin{pmatrix}
1 & \lambda(t) & & & \\
0 & 1 & & & \\
 & & 1 & & \\
 & & & \ddots & \\
 & & & & 1
\end{pmatrix}
\begin{pmatrix}
\beta_1(t) & & & & & & \\
 & \beta_2(t) & & & & & \\
 & & \ddots & & & & \\
 & & & \beta_r(t) & & & \\
 & & & & 0 & & \\
 & & & & & \ddots & \\
 & & & & & & 0
\end{pmatrix}
\begin{pmatrix}
\eta_1(t) & \eta_2(t) & & & \\
\zeta_1(t) & \zeta_2(t) & & & \\
 & & 1 & & \\
 & & & \ddots & \\
 & & & & 1
\end{pmatrix}
$$

$$
= \begin{pmatrix}
\delta(t) & 0 & & & & & \\
\beta_2(t)\zeta_1(t) & \beta_2(t)\zeta_2(t) & & & & & \\
 & & \beta_3(t) & & & & \\
 & & & \ddots & & & \\
 & & & & \beta_r(t) & & \\
 & & & & & 0 & \\
 & & & & & & \ddots \\
 & & & & & & & 0
\end{pmatrix}
$$

is an associate of $A(t)$. Since the degree $\sigma\delta(t) < \sigma\beta_1(t)$, it follows from the usual proof of Lemma 3 that there exists an associate $C(t)$ of the form (23) with $\sigma\gamma_1(t) < \sigma\beta_1(t)$. We repeat the argument with $C(t)$ in place of $B(t)$. After a finite number of repetitions of this process we arrive at an associate of $A(t)$ of the form (23) in which the first element has a bound which divides the succeeding elements. The theorem then follows from (1).

An associate (23) satisfying the conditions of Theorem 4 is a *normal form* for $A(t)$. The polynomials $\beta_i(t)$ may be taken to have leading coefficients $= 1$. When this has been done the $\beta_i(t) \neq 1$ are called the *invariant factors* of the normal form.

6. **The invariant factors of a p.l.t.** We return to the discussion of the p.l.t. **T** whose matrix T relative to the basis $(e_1, \cdots, e_n)$ of $\mathfrak{R}$ is given by (6).

Let $\mathfrak{R}[t]$ be the extension of $\mathfrak{R}$ obtained by allowing the coefficients to range in $\mathfrak{F}[t]$: $\mathfrak{R}[t]$ is the totality of forms $x(t) = \sum e_i \xi_i(t)$, $\xi_i(t) \in \mathfrak{F}[t]$. This definition is evidently independent of the choice of basis of $\mathfrak{R}$. $\mathfrak{R}[t]$ is an abelian group under addition. The correspondence $x(t) \to x(t)\alpha(t)$, $\alpha(t) \in \mathfrak{F}[t]$ is an automorphism of this group. Thus $\mathfrak{R}[t]$ may be looked upon as an abelian group with operators $\mathfrak{F}[t]$, $(\mathfrak{R}[t], \mathfrak{F}[t])$.

---

[13] Wedderburn [17] p. 134.

With $x(t) = \sum e_i\xi_i(t)$ we associate the vector $x = \sum e_i\xi_i(\mathbf{T})$ and with the automorphism $x(t) \to x(t)\alpha(t)$ we associate the automorphism $x \to x\alpha(\mathbf{T})$. These correspondences define an operator homomorphism of $(\mathfrak{R}[t], \mathfrak{F}[t])$ into $(\mathfrak{R}, \mathfrak{F}[\mathbf{T}])$.[14] Let $\mathfrak{N}$ denote the allowable subgroup sent into 0 by this homomorphism and define $f_1(t), \cdots, f_n(t)$ by

$$(f_1(t), \cdots, f_n(t)) = (e_1, \cdots, e_n)(T - t1) \tag{25}$$

where 1 denotes the unit matrix.

LEMMA 4. *$(f_1(t), \cdots, f_n(t))$ is an independent basis for $\mathfrak{N}$.*

The correspondent of $f_j(t)$ in the homomorphism of $(\mathfrak{R}[t], \mathfrak{F}[t])$ and $(\mathfrak{R}, \mathfrak{F}[\mathbf{T}])$ is

$$e_1\tau_{1j} + \cdots + e_{j-1}\tau_{j-1,j} + e_j(\tau_{jj} - \mathbf{T}) + e_{j+1}\tau_{j+11,j} + \cdots + e_n\tau_{nj} = 0$$

because of (6). Thus $f_j(t) \in \mathfrak{N}$. If $\sum f_j(t)\delta_j(t) = 0$, then

$$t\delta_j(t) = \sum \tau_{ji}\delta_i(t) \qquad (j = 1, \cdots, n) \tag{26}$$

since the e's are independent. If any $\delta(t) \neq 0$, let $\delta_s(t)$ be one of maximum degree. The equation (26) for $j = s$ is impossible. Hence all $\delta(t)$ must be 0 and the $f_j(t)$ are independent. Now let $x(t) = \sum e_i\xi_i(t)$ be any element of $\mathfrak{R}[t]$. By means of the relations

$$e_jt = -f_j(t) + \sum e_i\tau_{ij} \qquad (j = 1, \cdots, n) \tag{25'}$$

we may express $x(t)$ in the form

$$x(t) = \sum f_i(t)\psi_i(t) + \sum e_i\xi_i$$

where $\xi_i \in \mathfrak{F}$. If $x(t) \in \mathfrak{N}$, we have $x = \sum e_i\xi_i(\mathbf{T}) = 0 = \sum e_i\xi_i$ since $f_i(\mathbf{T}) = 0$. Thus $\xi_i = 0$ and $x(t) = \sum f_i(t)\psi_i(t)$ as was to be shown.

We may replace the bases $(e_1, \cdots, e_n)$ and $(f_1(t), \cdots, f_n(t))$ of $\mathfrak{R}[t]$ and $\mathfrak{N}$ respectively by

$$(e_1^*(t), \cdots, e_n^*(t)) = (e_1, \cdots, e_n)U^{-1}(t),$$
$$(f_1^*(t), \cdots, f_n^*(t)) = (f_1(t), \cdots, f_n(t))V(t),$$

where $V(t)$, $U^{-1}(t)$ are units in $\mathfrak{F}[t]_n$. Then (25) becomes

$$(f_1^*(t), \cdots, f_n^*(t)) = (e_1^*(t), \cdots, e_n^*(t))T^*(t)$$

where

$$T^*(t) = U(t)(T - t1)V(t).$$

In view of §5 we may choose $U(t)$ and $V(t)$ so that $T^*(t)$ has the form

$$T^*(t) = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & \mu_1(t) & & \\ & & & & \ddots & \\ & & & & & \mu_r(t) \end{pmatrix}$$

[14] This is essentially the method used by Krull [9] pp. 22–32 to obtain the canonical form of ordinary linear transformations.

where the invariant factors $\mu_i(t)$ are bounded with bounds $\mu_i^*(t)$ dividing $\mu_j(t)$ if $i < r, < j$. We shall call the set $\{\mu_k(t)\}$ $(k = 1, \cdots, r)$ a set of *invariant factors of the p.l.t.* $\mathbf{T}$.

Referring back to $(\mathfrak{R}, \mathfrak{F}[\mathbf{T}])$ which is operator isomorphic to the factor group $(\mathfrak{R}[t], \mathfrak{F}[t])/\mathfrak{N}$, we may conclude that the vectors $e_1^*, \cdots, e_n^*$ corresponding to $e_1^*(t), \cdots, e_n^*(t)$ generate $(\mathfrak{R}, \mathfrak{F}[\mathbf{T}])$ in the sense each of its vectors may be written as $\sum e_i^* \xi_i(\mathbf{T})$. Since $e_1^*(t), \cdots, e_{n-r}^*(t) \in \mathfrak{N}$, $e_1^* = 0, \cdots, e_{n-r}^* = 0$ and so may be discarded. We denote e_{n-r+k}^* by g_k $(k = 1, \cdots, r)$. If $\sum g_k \delta_k(\mathbf{T}) = 0$, we have, since

$$f_1^*(t) = e_1^*(t), \cdots, f_{n-r}^*(t) = e_{n-r}^*(t), f_{n-r+1}^*(t) = e_{n-r+1}^*(t)\mu_1(t), \cdots, f_n^*(t) = e_n^*(t)\mu_r(t)$$

form a basis for $\mathfrak{N}$, that $\delta_k(t)$ is a r.m. of $\mu_k(t)$. It follows that if

$$\mu_k(t) = t^{n_k} - t^{n_k-1}\tau_1^{(k)} - \cdots - \tau_{n_k}^{(k)}$$

then

$$(g_1, g_1\mathbf{T}, \cdots, g_1\mathbf{T}^{n_1-1}; \cdots; g_r, g_r\mathbf{T}, \cdots, g_r\mathbf{T}^{n_r-1})$$

is an ordinary independent basis for $\mathfrak{R}$. Thus $\sum n_i = n$. Relative to this basis T the matrix of $\mathbf{T}$ has the form

$$\begin{pmatrix} T_1 & & & \\ & T_2 & & \\ & & \ddots & \\ & & & T_r \end{pmatrix} \tag{27}$$

where

$$T_k = \begin{pmatrix} 0 & \cdot & \cdot & \cdot & \tau_{n_k}^{(k)} \\ 1 & 0 & & & \cdot \\ & \cdot & \cdot & & \cdot \\ & & \cdot & \cdot & \cdot \\ & & & 1 & \tau_1^{(k)} \end{pmatrix} \tag{28}$$

The invariant subspace $\mathfrak{R}_k = (g_k, g_k\mathbf{T}, \cdots)$ is generated by the single vector g_k (acted upon by $\mathfrak{F}[\mathbf{T}]$). Let $\mathbf{T}_k$ denote the contraction of $\mathbf{T}$ in $\mathfrak{R}$. Thus if $g \in \mathfrak{R}_k$, $g\mathbf{T}_k \equiv g\mathbf{T}$. We shall call $\mathfrak{R}_k$ a *cyclic subspace* and $\mathbf{T}_k$ a *cyclic p.l.t.* The polynomial $\mu_k(t)$ is the *order* of the *generator* g_k of $\mathfrak{R}_k$. The above result may be stated as

Theorem 5. $\mathbf{T} = \mathbf{T}_1 \oplus \mathbf{T}_2 \oplus \cdots \oplus \mathbf{T}_r$ *where the* $\mathbf{T}_k$ *are cyclic with generators* g_k *whose orders* $\mu_k(t)$ *are the invariant factors of a normal form of the matrix* $T - t1$.

7. **Cyclic p.l.t.** In this section we suppose that $\mathbf{T}$ acting in $\mathfrak{R}$ is cyclic and that g is a generator of $\mathfrak{R}$ whose order is $\mu(t)$. $\sigma\mu(t) = n$. If $h = g\alpha(\mathbf{T})$, its order is readily seen to be $\nu(t) = \alpha(t)^{-1}\cdot\mu(t)\cdot\alpha(t)$. Thus h is a generator of $\mathfrak{R}$ if and only if $\sigma\nu(t) = \sigma\mu(t)$, i.e. $\mu(t) \simeq \nu(t)$. Hence

LEMMA 5. *A n.a.s.c. that two cyclic p.l.t. be similar is that the orders of their corresponding generators be similar polynomials.*[15]

The vectors of $\mathfrak{R}$ may be represented uniquely in the form $g\lambda(\mathbf{T})$ where $\lambda(t)$ is of degree $<n$. Now let $\mathfrak{R}_1$ be an invariant subspace of $\mathfrak{R}$ and $\mathbf{T}_1$ the contraction of $\mathbf{T}$ in $\mathfrak{R}_1$. Let $g_1 = g\lambda_1(\mathbf{T}_1)$ be a vector of $\mathfrak{R}_1$ for which the polynomial $\lambda_1(t)$ has minimum degree for the vectors of $\mathfrak{R}_1$. It is an immediate consequence of the division process that every vector of $\mathfrak{R}_1$ is a "multiple" of g_1, i.e. $\mathbf{T}_1$ is cyclic with generator g_1. Further, the order $\mu(t)$ is a r.m. of $\lambda_1(t)$, $\mu(t) = \lambda_1(t)\mu_1(t)$. The right associate of $\mu_1(t)$ whose leading coefficient is 1 is the order of g_1.

LEMMA 6. *A n.a.s.c. that the cyclic* $\mathbf{T}$ *be irreducible is that the order* $\mu(t)$ *be an irreducible polynomial.*

The above results establish a correspondence between a factorization of $\mu(t)$ into irreducible factors and a composition series of $\mathbf{T}$. From this correspondence it is easily seen that Ore's theorem on the unique factorization (in the sense of similarity) of polynomials in $\mathfrak{F}[t]$ is precisely the Jordan-Hölder theorem applied to cyclic p.l.t. A similar connection exists between the concepts of decomposability applied to polynomials and to p.l.t.

Suppose that the order $\mu(t)$ is decomposable, say

$$\mu(t) = \lambda_1(t)\mu_1(t) = \lambda_2(t)\mu_2(t)$$

where

$$(\lambda_1(t), \lambda_2(t))_L = 1 = (\mu_1(t), \mu_2(t))_R$$

and all of these polynomials are normalized. Set $g_1 = g\lambda_1(\mathbf{T})$, $g_2 = g\lambda_2(\mathbf{T})$ and let $\mathfrak{R}_1$ and $\mathfrak{R}_2$ be the cyclic subspaces generated by these vectors. We may determine $\nu_1(t)$, $\nu_2(t)$ so that

$$\lambda_1(t)\nu_1(t) + \lambda_2(t)\nu_2(t) = 1$$

and hence

$$g = g_1\nu_1(\mathbf{T}) + g_2\nu_2(\mathbf{T}).$$

By comparing the dimensions of $\mathfrak{R}$, $\mathfrak{R}_1$, and $\mathfrak{R}_2$ we obtain $\mathfrak{R}_1 \cap \mathfrak{R}_2 = 0$ and hence $\mathbf{T} = \mathbf{T}_1 \oplus \mathbf{T}_2$. Conversely, if $\mathbf{T} = \mathbf{T}_1 \oplus \mathbf{T}_2$ and correspondingly $\mathfrak{R} = \mathfrak{R}_1 \oplus \mathfrak{R}_2$, let $g_1 = g\lambda_1(\mathbf{T})$, $g_2 = g\lambda_2(\mathbf{T})$ be generators of $\mathfrak{R}_1$ and $\mathfrak{R}_2$ such that $\lambda_1(t)$ and $\lambda_2(t)$ are normalized l.f.'s of $\mu(t)$. The orders $\mu_1(t)$, $\mu_2(t)$ of g_1 and g_2 are then normalized r.f.'s of $\mu(t)$. Since $g = g_1\nu_1(\mathbf{T}) + g_2\nu_2(\mathbf{T})$, we have

$$(\lambda_1(t), \lambda_2(t))_L = 1, \qquad [\mu_1(t), \mu_2(t)]_L = \mu(t).$$

Since $\mathfrak{R}_1 \cap \mathfrak{R}_2 = 0$,

$$\sigma\mu(t) = \sigma\mu_1(t) + \sigma\mu_2(t)$$

and hence

$$(\mu_1(t), \mu_2(t))_R = 1.$$

Thus we have proved

[15] Noether-Schmeidler [10], p. 21.

LEMMA 7. *A n.a.s.c. that a cyclic* **T** *be indecomposable is that* $\mu(t)$ *be an indecomposable polynomial*

Lemma 6 and 7 give

LEMMA 8. *A n.a.s.c. that a cyclic* **T** *be completely reducible is that* $\mu(t)$ *be a completely reducible polynomial.*

8. **Criteria for similarity, reducibility, etc.** Theorem 5 enables us to apply the lemmas of the preceding section to arbitrary p.l.t. For according to this theorem a p.l.t. **T** is decomposable unless it is cyclic and the condition that it be cyclic is that the normal form of $T - t1$ have but one invariant factor, i.e. $r = 1$. Thus Lemma 6 and 7 give respectively

THEOREM 6. *A n.a.s.c. that* **T** *be irreducible is that it be cyclic and its order* $\mu(t)$ *be an irreducible polynomial.*

THEOREM 7. *A n.a.s.c. that* **T** *be indecomposable is that it be cyclic and* $\mu(t)$ *be an indecomposable polynomial.*

The condition that **T** be completely reducible is, according to Lemma 8 that all of the invariant factors $\mu_k(t)$ be completely reducible. However, if we recall that every factor of a completely reducible polynomial is completely reducible, we obtain

THEOREM 8. *A n.a.s.c. that* **T** *be completely reducible is that the last invariant factor* $\mu_r(t)$ *be a completely reducible polynomial.*

Let

$$\mathbf{T} = \mathbf{T}_1 \oplus \mathbf{T}_2 \oplus \cdots \oplus \mathbf{T}_u \tag{29}$$

be a decomposition of **T** into indecomposable parts. By Theorem 3 each $\mathbf{T}_i$ is cyclic and the order $\rho_i(t)$ of its generator is indecomposable. $\rho_i(t)$ is uniquely determined in the sense of similarity by $\mathbf{T}_i$. We shall call these polynomials the *elementary divisors of* **T**. In virtue of the theorem of Krull that the indecomposable parts of an abelian group with operators are unique in the sense of isomorphism, we have that the $\mathbf{T}_i$ in (29) are unique in the sense of similarity and hence the elementary divisors are determined in the same sense by **T**. Hence

THEOREM 9. *A n.a.s.c. that two p.l.t. be similar is that their elementary divisors be similar in pairs.*

The following theorem gives some additional information regarding the invariance of the invariant factors $\mu_i(t)$.

THEOREM 10. $\mathbf{T}_r$ *is a cyclic contraction of maximum dimension* n_r *of* **T**. *If* $\lambda(t)$ *is the order of any cyclic contraction of dimension* n_r, *then* $\lambda(t) = \lambda^{(1)}(t)\mu_{r-1}^*(t)$, $\mu_r(t) = \mu^{(1)}(t)\mu_{r-1}^*(t)$ *where* $\mu_{r-1}^*(t)$ *is the bound of* $\mu_{r-1}(t)$ *and* $\lambda^{(1)}(t) \simeq \mu^{(1)}(t)$.

In the notation of Theorem 5 suppose $\mathbf{T} = \mathbf{T}_1 \oplus \mathbf{T}_2 \oplus \cdots \oplus \mathbf{T}_r$ where $\mathbf{T}_i$ has generator g_i of order $\mu_i(t)$. Let $\lambda(t)$ denote the order of

$$g = g_1\alpha_1(\mathbf{T}) + \cdots + g_r\alpha_r(\mathbf{T}).$$

The order of $g_i\alpha_i(\mathbf{T})$ is $\lambda_i(t) = \alpha_i(t)^{-1}\cdot\mu_i(t)\cdot\alpha_i(t)$ and $\lambda(t) = [\lambda_1(t), \cdots, \lambda_r(t)]_R$. Since $[\lambda_1(t), \cdots, \lambda_{r-1}(t)]_R$ is a factor of $\mu_{r-1}^*(t)$, $\lambda(t)$ is a l.f. of $[\mu_{r-1}^*(t), \lambda_r(t)]_R$.

Since $\mu_r(t) = \mu^{(1)}(t)\mu^*_{r-1}(t)$, $\lambda_r(t)$ is a l.f. of $\lambda^{(1)}(t)\mu^*_{r-1}(t)$ where $\lambda^{(1)}(t) = \alpha_r(t)^{-1}\cdot\mu^{(1)}(t)\cdot\alpha_r(t)$. Thus $\lambda(t)$ is a l.f. of $\lambda^{(1)}(t)\mu^*_{r-1}(t)$. Since $\sigma\lambda^{(1)}(t) \leqq \sigma\mu^{(1)}(t)$ we have $\sigma\lambda(t) \leqq \sigma\mu_r(t)$. If equality is to hold we must have $\sigma\lambda^{(1)}(t) = \sigma\mu^{(1)}(t)$, i.e. $\lambda^{(1)}(t) \simeq \mu^{(1)}(t)$.

Special cases.

(1). If **T** is an l.t. and $\mathfrak{F}$ is commutative, $\mathfrak{F}[t]$ is commutative and similarity implies identity. Theorem 9 shows that the elementary divisors are a complete set of invariants of **T** and Theorem 10 may be extended to give the same result for the invariant factors.

(2). If **T** is an s.l.t. with automorphism of relative infinite order, the finite elements of $\mathfrak{F}[t]$ consist of the powers t^m. Hence the bounded elements are also of this form. Thus in the decomposition $\mathbf{T} = \mathbf{T}_1 \oplus \cdots \oplus \mathbf{T}_{r-1} \oplus \mathbf{T}_r$ of Theorem 5 $\mathbf{T}_1 \oplus \cdots \oplus \mathbf{T}_{r-1}$ is nilpotent and $\mathbf{T}_r$ is cyclic.

(3). If **T** is a d.t. and $\mathfrak{F}$ has characteristic 0, 1 is the only normalized bounded element in $\mathfrak{F}[t]$. Thus $r = 1$ and **T** is cyclic. This result gives the well-known equivalence between the theory of systems of n linear homogeneous differential equations (analytic case) and that of a single n-th order linear homogeneous differential equation.

9. **Finite p.l.t.** The homomorphism $\mathfrak{F}[t] \sim \mathfrak{F}[\mathbf{T}]$ between the polynomial ring and the ring of a p.l.t. defines an isomorphism $\mathfrak{F}[\mathbf{T}] \simeq \mathfrak{F}[t]/\mathfrak{J}$ where $\mathfrak{J}$ is a two-sided ideal. If $\mathbf{T} \neq 0$, then $\mathfrak{J} \neq (1)$. Thus either $\mathfrak{J} = (0)$ or $\mathfrak{J} = (\mu^*(t))$ where $\mu^*(t)$ is a normalized finite polynomial $\neq 0, \neq 1$. In the latter case we shall call **T** *finite* and $\mu^*(t)$ its *minimum function.* $\mu^*(t)$ may be defined also as the normalized polynomial of least degree such that $x\mu^*(\mathbf{T}) = 0$ for all x in $\mathfrak{R}$.

Let **T** be finite and decomposed as $\mathbf{T}_1 \oplus \cdots \oplus \mathbf{T}_r$ into the cyclic parts $\mathbf{T}_i$ with generators g_i whose orders $\mu_i(t)$ are chosen as in Theorem 5. We have $g_r\alpha(\mathbf{T})\mu^*(\mathbf{T}) = 0$ and hence $\mu^*(t)$ is a r.m. of $\alpha(t)^{-1}\cdot\mu_r(t)\cdot\alpha(t)$. Since $\alpha(t)$ is arbitrary, it follows that $\mu_r(t)$ is bounded with bound $\mu_r^*(t)$ a factor of $\mu^*(t)$. On the other hand it is easily seen that $x\mu_r^*(\mathbf{T}) = 0$ for all x. Thus $\mu_r^*(t) = \mu^*(t)$.

THEOREM 11. *A n.a.s.c. that* **T** *be finite is that $\mu_r(t)$ be bounded. If the condition is satisfied, the minimum function of* **T** *is $\mu_r^*(t)$.*

Suppose $\mu^*(t) = \pi_1^*(t)^{e_1}\pi_2^*(t)^{e_2}\cdots\pi_m^*(t)^{e_m}$ where $\pi_i^*(t)$ are distinct and irreducible in $\mathfrak{B}^*$. Set $\pi_i^*(t)^{e_i} = \rho_i^*(t)$ and $\mu^*(t)\rho_i^*(t)^{-1} = \tau_i^*(t)$. Let $\mathfrak{R}_i^*$ denote the subspace of $\mathfrak{R}$ consisting of all vectors of the form $x\tau_i^*(\mathbf{T})$. Since $\tau_i^*(t)$ is finite $\mathfrak{R}_i^*$ is an invariant subspace. If $x_i \in \mathfrak{R}_i^*$ then $x_i\rho_i^*(\mathbf{T}) = 0$ and since $(\rho_i^*(t), \tau_i^*(t)) = 1$,

$$\mathfrak{R}_i^* \cap (\mathfrak{R}_1^* \cup \cdots \cup \mathfrak{R}_{i-1}^* \cup \mathfrak{R}_{i+1}^* \cup \cdots \mathfrak{R}_m^*) = 0.$$

Since

$$(\tau_1^*(t), \cdots, \tau_m^*(t)) = 1,$$

$$\mathfrak{R} = \mathfrak{R}_1^* \cup \cdots \cup \mathfrak{R}_m^*$$

and correspondingly

$$\mathbf{T} = \mathbf{T}_1 \oplus \mathbf{T}_2 \oplus \cdots \oplus \mathbf{T}_m.$$

Thus we have proved

THEOREM 12. *If* $\mathbf{T}$ *is finite and indecomposable, then* $\mu^*(t)$ *is a power of a polynomial irreducible in* $\mathfrak{B}^*$.

The following is an extension of a well-known result on ordinary linear transformations.

THEOREM 13. *A n.a.s.c. that a finite* $\mathbf{T}$ *be completely reducible is that its minimum function have no multiple factors in* $\mathfrak{B}^*$.

If $\mathbf{T}$ is completely reducible, its elementary divisors are irreducible polynomials and hence their bounds are irreducible in $\mathfrak{B}^*$. It is readily seen that $\mu^*(t)$ is the l.c.m. of these bounds and hence has no multiple factors in $\mathfrak{B}^*$. To prove the converse we note that if $\pi(t)$ is an elementary divisor of $\mathbf{T}$, its bound $\pi^*(t)$ is a factor of $\mu^*(t)$ and hence is irreducible in $\mathfrak{B}^*$. Since $\pi^*(t)$ can be defined as a l.c.m. of irreducible polynomials, it is completely reducible. Hence its factor $\pi(t)$ is also completely reducible, but being indecomposable, it must be irreducible. Thus every elementary divisor of $\mathbf{T}$ is irreducible and so $\mathbf{T}$ is completely reducible.[16]

LEMMA 9. *If* $\rho(t)$ *is a normalized indecomposable bounded polynomial and* $\rho^*(t)$ *its bound, then the indecomposable parts of* $\rho^*(t)$ *are similar to* $\rho(t)$.

Suppose $\lambda_k(t)$ is a normalized polynomial decomposable into k parts similar to $\rho(t)$. $\lambda_k(t)$ is a factor of $\rho^*(t)$. Suppose $\lambda_k(t) \neq \rho^*(t)$ and consider the p.l.t. $\mathbf{T}$ defined as the direct sum of (decomposable into) the two cyclic p.l.t.'s whose orders are $\rho(t)$ and $\lambda_k(t)$. Thus $\mathbf{T}$ has a matrix of the form

$$T = \begin{pmatrix} R & 0 \\ 0 & L_k \end{pmatrix}$$

where R and L_k are associated with $\rho(t)$ and $\lambda_k(t)$ as in (28). We assert that $\mathbf{T}$ is cyclic. For otherwise the number r of invariant factors $\mu_i(t)$ in Theorem 5 is >1. By Theorem 10 $\mu_r(t)$ is the order of a cyclic subspace of maximum dimension of $\mathbf{T}$ and if $\sigma\mu_r(t) = \sigma\lambda_k(t)$, $\lambda_k(t)$ is divisible by $\mu^*_{r-1}(t)$. We obtain a decomposition of $\mathbf{T}$ into indecomposable parts by decomposing the $\mu_i(t)$. Thus $\mu_{r-1}(t)$ is divisible by a polynomial similar to $\rho(t)$ and so $\lambda_k(t)$ is divisible by $\rho^*(t)$ contrary to the assumption that $\rho^*(t) \neq \lambda_k(t)$. On the other hand if $\sigma\mu_r(t) > \sigma\lambda_k(t)$ we must have $\sigma\mu_{r-1}(t) < \sigma\rho(t)$ and hence $\mathbf{T}$ has indecomposable parts not similar to $\rho(t)$ contrary to Krull's theorem. Thus $r > 1$ is impossible and $\mathbf{T}$ is cyclic. Let $\lambda_{k+1}(t)$ denote its order. $\lambda_{k+1}(t)$ is decomposable into $k+1$ parts similar to $\rho(t)$ and hence divides $\rho^*(t)$. If $\lambda_{k+1}(t) \neq \rho^*(t)$ we repeat the process. If we start with $\lambda_1(t) = \rho(t)$ we arrive in this way after a finite

[16] The remainder of this section was added Aug. 14, 1936. These results are related in part to some recent work by Prof. M. H. Ingraham and Dr. M. C. Wolf on matrices with elements in a division algebra. I am indebted to Prof. Ingraham for communicating a copy of their manuscript to me.

number of steps at $\rho^*(t)$ and obtain a decomposition for it into parts similar to $\rho(t)$.

An immediate consequence of Lemma 9 and Krull's theorem is

LEMMA 10. *If $\rho_1(t)$ and $\rho_2(t)$ are indecomposable bounded polynomials with the same bound, then $\rho_1(t) \simeq \rho_2(t)$.*

From this and Theorem 9 we have

THEOREM 14. *A n.a.s.c. that two finite p.l.t. be similar is that the bounds of their elementary divisors be identical.*

Let $\pi(t)$ be any irreducible bounded polynomial with bound $\pi^*(t)$ and $\rho_e(t) = \pi_1(t) \cdots \pi_e(t)$ where the $\pi_i(t) \simeq \pi(t)$. Each $\pi_i(t)$ divides $\pi^*(t)$, say $\pi^*(t) = \pi_i(t)\gamma_i(t) = \delta_i(t)\pi_i(t)$. Suppose

$$\pi^*(t)^{e-1} = \rho_{e-1}(t)\alpha(t) \qquad \rho_{e-1}(t) = \pi_1(t) \cdots \pi_{e-1}(t).$$

Then

$$\pi^*(t)^e = \rho_{e-1}(t)\alpha(t)\pi^*(t) = \rho_{e-1}(t)\pi^*(t)\beta(t) = \rho_e(t)\gamma_e(t)\beta(t).$$

Hence $\rho_e^*(t)$ the bound of $\rho_e(t)$ is $\pi^*(t)^f$ where $f \leqq e$. We shall show next that $\pi_1(t), \cdots, \pi_e(t)$ may be so chosen that $\rho_e^*(t) = \pi^*(t)^e$. Suppose $\rho_{e-1}(t)$ has already been found so that $\rho_{e-1}^*(t) = \pi^*(t)^{e-1}$. There exists an irreducible $\pi_e(t) \simeq \pi(t)$ such that the bound of $\rho_e(t) = \rho_{e-1}(t)\pi_e(t)$ is $\pi^*(t)^e$. For otherwise all the $\rho_e(t)$ divide $\pi^*(t)^{e-1}$ and hence their l.c.r.m. $\rho_{e-1}(t)\pi^*(t)$ divides $\pi^*(t)^{e-1}$. But then $\rho_{e-1}(t)$ divides $\pi^*(t)^{e-2}$ contrary to assumption. We require these remarks to prove the following extensions of Theorem 12.

THEOREM 15. *A n.a.s.c. that the bounded $\rho_e(t) = \pi_1(t) \cdots \pi_e(t)$ ($\pi_i(t)$ irreducible and $\simeq$ to $\pi(t)$) be indecomposable is that its bound $\rho_e^*(t) = \pi^*(t)^e$.*

Sufficiency. If $\rho_e(t) = [\rho_{e_1}^{(1)}(t), \rho_{e_2}^{(2)}(t)]_R$ where $\rho_{e_i}^{(i)}(t) = \pi_1^{(i)}(t) \cdots \pi_{e_i}^{(i)}(t)$ and $\pi_j^{(i)}(t) \simeq \pi(t)$ then $\rho_{e_i}^{(i)}(t)$ divides $\pi^*(t)^{e_i}$ and hence $\rho_e(t)$ divides $\pi^*(t)^f$ where $f = \max(e_1, e_2)$. Hence we can not have both e_1 and $e_2 < e$.

Necessity. Suppose $\rho_e^*(t) = \pi^*(t)^f$ where $f \leqq e$. We have seen that a polynomial $\rho_f'(t) = \pi_1'(t) \cdots \pi_f'(t)$ exists having the bound $\pi^*(t)^f$. It follows that $\rho_f'(t)$ is indecomposable and by Lemma 10 is $\simeq \rho_e(t)$. Thus $e = f$.

THEOREM 16. *A n.a.s.c. that a finite polynomial be indecomposable is that it be divisible on the left (right) by only one normalized irreducible polynomial.*

The condition is evidently sufficient. Now if $\rho_e(t)$ is indecomposable and divisible by the distinct normalized irreducible polynomials $\pi_1(t)$ and $\pi_2(t)$, then $\rho_e(t)$ is divisible by their l.c.r.m. $\mu(t)$, i.e. $\rho_e(t) = \mu(t)\pi_3(t) \cdots \pi_e(t)$. Since $\mu(t)$ divides $\pi^*(t)$ it follows easily that $\rho_e(t)$ divides $\pi^*(t)^{e-1}$ contrary to Theorem 15.

As a consequence of this theorem we note that if $\rho_e(t) = \pi_1(t) \cdots \pi_e(t)$ is indecomposable, so is $\rho_k(t) = \pi_1(t) \cdots \pi_k(t)$, $1 \leqq k \leqq e$.

If $x(\mathbf{T})$ is a variable element of $\mathfrak{F}[\mathbf{T}]$, the correspondence $x(\mathbf{T}) \rightarrow x(\mathbf{T})\mathbf{T}$ is a p.l.t. $\tilde{\mathbf{T}}$ determined by $\mathbf{T}$ in the vector space $\mathfrak{F}[\mathbf{T}]$. The correspondence $\mathbf{T} \rightarrow \tilde{\mathbf{T}}$ defines an isomorphism between $\mathfrak{F}[\mathbf{T}]$ and $\mathfrak{F}[\tilde{\mathbf{T}}]$ analogous to the regular representation of a hypercomplex system. By means of this isomorphism Theorems 12 and 13 yield results on the structure of $\mathfrak{F}[\mathbf{T}]$. Thus it follows that a n.a.s.c.

that $\mathfrak{F}[\mathbf{T}]$ be a semi-simple ring is that $\mu^*(t)$ have no multiple factors in $\mathfrak{P}^*$. This will also be a consequence of a general theorem of the next section.

10. **The automorphism ring of a p.l.t.** An (*operator*) *automorphism* **A** of $(\mathfrak{R}, \mathfrak{F}[\mathbf{T}])$ (or of **T**) is defined by the conditions

$$(x + y)\mathbf{A} = x\mathbf{A} + y\mathbf{A} \tag{30a}$$

$$(x\mathbf{A})\mathbf{O} = (x\mathbf{O})\mathbf{A} \tag{30b}$$

for every **O** in $\mathfrak{F}[\mathbf{T}]$. It is sufficient to require

$$(x\mathbf{A})\alpha = (x\alpha)\mathbf{A} \qquad (x\mathbf{A})\mathbf{T} = (x\mathbf{T})\mathbf{A}$$

in place of (30b). Thus **A** is an l.t. commutative with **T** and conversely.

Let **A** be an arbitrary l.t., **T** a p.l.t. and x a vector such that $(x\mathbf{A})\mathbf{T} = (x\mathbf{T})\mathbf{A}$. Then

$$(x\alpha)\mathbf{TA} = (x\mathbf{TA})\bar{\alpha} + (x\mathbf{A})\alpha' \qquad (x\alpha)\mathbf{AT} = (x\mathbf{AT})\bar{\alpha} + (x\mathbf{A})\alpha'$$

so that $(x\alpha)\mathbf{TA} = (x\alpha)\mathbf{AT}$. Thus the totality of vectors for which **A** and **T** commute constitutes a vector subspace of $\mathfrak{R}$. Hence **A** is an automorphism of **T** is and only if $e_k\mathbf{TA} = e_k\mathbf{AT}$ $(k = 1, \cdots, n)$ where $(e_1, \cdots, e_n)$ is a basis for $\mathfrak{R}$. This condition in terms of the matrices A, T of **A** and **T** is

$$AT = T\bar{A} + A'. \tag{31}$$

In a similar fashion we define an *automorphism* $\mathbf{A}(t)$ of $(\mathfrak{R}[t], \mathfrak{F}[t])$ by

$$[x(t) + y(t)]\mathbf{A}(t) = x(t)\mathbf{A}(t) + y(t)\mathbf{A}(t)$$

$$[x(t)\alpha(t)]\mathbf{A}(t) = [x(t)\mathbf{A}(t)]\alpha(t).$$

If $(e_1(t), \cdots, e_n(t))$ is a basis of $\mathfrak{R}[t]$ then $\mathbf{A}(t)$ is determined by the matrix $A(t)$ given by

$$(e_1(t)\mathbf{A}(t), \cdots, e_n(t)\mathbf{A}(t)) = (e_1(t), \cdots, e_n(t))A(t)$$

and conversely any matrix defines an automorphism. If $\mathbf{A}(t)$ leaves $\mathfrak{N}$ invariant where $(\mathfrak{R}, \mathfrak{F}[\mathbf{T}]) \simeq (\mathfrak{R}[t], \mathfrak{F}[t])/\mathfrak{N}$, then $\mathbf{A}(t)$ induces an automorphism in the quotient group $(\mathfrak{R}, \mathfrak{F}[\mathbf{T}])$, the *projection* of $\mathbf{A}(t)$. Suppose

$$(f_1(t), \cdots, f_n(t)) = (e_1(t), \cdots, e_n(t))T(t)$$

is a basis of $\mathfrak{N}$. We recall that $T(t)$ is an associate of $T - t1$. The condition that $A(t)$ leave $\mathfrak{N}$ invariant is

$$A(t)T(t) = T(t)A_1(t). \tag{32}$$

The projection **A** of $\mathbf{A}(t)$ is 0 if and only if

$$A(t) \equiv 0 \qquad (T(t))_R$$

where $(T(t))_R$ denotes the ideal of right multiples of $T(t)$.

Conversely if $\mathbf{A}$ is any automorphism of $(\mathfrak{R}, \mathfrak{F}[\mathbf{T}])$, it may be extended to an automorphism $\mathbf{A}(t)$ of $(\mathfrak{R}[t], \mathfrak{F}[t])$. For if $x(t) = e_1\xi_1(t) + \cdots + e_n\xi_n(t)$, then we may define $A(t)$ by

$$x(t)\mathbf{A}(t) = e_1\mathbf{A}\xi_1(t) + \cdots + e_n\mathbf{A}\xi_n(t).$$

This is independent of the choice of the basis $(e_1, \cdots, e_n)$. Now if $x(t) \in \mathfrak{N}$. i.e. $\sum e_i\xi_i(\mathbf{T}) = 0$, then since $\mathbf{A}$ commutes with all $\xi_i(\mathbf{T})$, $x(t)\mathbf{A}(t)$ also $\in \mathfrak{N}$. Thus A may be obtained as a projection of an $\mathbf{A}(t)$ leaving $\mathfrak{N}$ invariant.

Let $\mathfrak{A}(\mathbf{T}) = \mathfrak{A}$ denote the ring of automorphisms of $\mathbf{T}$. The above discussion shows that $\mathfrak{A}$ is inverse isomorphic to the ring of matrices with elements in $\mathfrak{F}$ satisfying (31) and also inverse isomorphic to the ring of matrices with elements in $\mathfrak{F}[t]$ satisfying (32) taken mod $(T(t))_R$. The latter will be called the *invariant ring* of $T(t)$.[17]

THEOREM 17. *If* $\mathbf{T}_1 \simeq \mathbf{T}_2$ *then* $\mathfrak{A}(\mathbf{T}_1) \simeq \mathfrak{A}(\mathbf{T}_2)$.

For $\mathbf{T}_2 = \mathbf{A}^{-1}\mathbf{T}_1\mathbf{A}$ and hence if $\mathbf{A}_1 \in \mathfrak{A}(\mathbf{T}_1)$, $\mathbf{A}_2 = \mathbf{A}^{-1}\mathbf{A}_1\mathbf{A} \in \mathfrak{A}(\mathbf{T}_2)$ and conversely.

THEOREM 18. *If* $\mathbf{T}$ *is irreducible,* $\mathfrak{A}(\mathbf{T})$ *is a field.*

If $\mathbf{A} \in \mathfrak{A}(\mathbf{T})$, the space $\mathfrak{R}\mathbf{A}$ consisting of all vectors $x\mathbf{A}$ is an invariant subspace of $\mathfrak{R}$. Hence either $\mathfrak{R}\mathbf{A} = 0$ or $\mathfrak{R}\mathbf{A} = \mathfrak{R}$, i.e. $\mathbf{A} = 0$ or $\mathbf{A}$ has an inverse $\mathbf{A}^{-1}$. Evidently $\mathbf{A}^{-1}$ also $\in \mathfrak{A}(\mathbf{T})$.

The following theorem, a special case of a theorem on the automorphisms of abelian groups due to v. d. Waerden, gives the structure of $\mathfrak{A}(\mathbf{T}) = \mathfrak{A}$ for a completely reducible p.l.t.

THEOREM 19. *Let*

$$\mathbf{T} = (\mathbf{T}_1^{(1)} \oplus \cdots \oplus \mathbf{T}_{n_1}^{(1)}) \oplus \cdots \oplus (\mathbf{T}_1^{(m)} \oplus \cdots \oplus \mathbf{T}_{n_m}^{(m)})$$

where each $\mathbf{T}_j^{(i)}$ *is irreducible and* $\mathbf{T}_k^{(i)} \simeq \mathbf{T}_l^{(i)}$ *but* $\mathbf{T}_k^{(i)} \not\simeq \mathbf{T}_l^{(j)}$ *if* $j \neq i$. *Then*

$$\mathfrak{A} = \mathfrak{A}^{(1)} \oplus \cdots \oplus \mathfrak{A}^{(m)}, \qquad \mathfrak{A}^{(i)} \simeq \mathfrak{D}_{n_i}^{(i)} \tag{33}$$

where $\mathfrak{D}^{(i)}$ *is a field, the automorphism ring of* $\mathbf{T}_j^{(i)}$.

For a proof of this theorem the reader is referred to v.d. Waerden [15] vol. II, pp. 165–168.

Consider the case of a finite completely reducible $\mathbf{T}$. Its minimum function $\mu^*(t) = \pi_1^*(t) \cdots \pi_m^*(t)$ where the $\pi_i^*(t)$ are irreducible in $\mathfrak{B}^*$ and $\pi_i^*(t) \neq \pi_j^*(t)$ $(i \neq j)$. Any two irreducible factors of $\pi_i^*(t)$ are similar. Let $\pi_i(t)$ be such a factor. Then $\pi_i(t) \not\simeq^1 \pi_j(t)$ and (33) gives the structure of $\mathfrak{A}$ where $\mathfrak{D}^{(i)}$ is inverse isomorphic to the invariant ring of $\pi_i(t)$. In particular if $m = 1$, i.e. $\mu^*(t)$ is irreducible in $\mathfrak{B}^*$, $\mathfrak{A}$ is a simple ring $\simeq \mathfrak{D}_{n_1}$.

11. Periodic s.l.t. and generalized cyclic algebras. An s.l.t. will be called *periodic* if some power of it, say $\mathbf{T}^m = \mu \neq 0$, i.e. $\mathbf{T}^m$ is the s.l.t. $x \to x\mu = x\mathbf{M}$

[17] Cf. Ore [11] p. 241. The connection between the ring of matrices commutative with a matrix in a commutative $\mathfrak{F}$ and the invariant ring was first noted by Frobenius. Cf. Wedderburn [18] p. 106.

whose automorphism is $\alpha \to \mu^{-1}\alpha\mu$ and matrix (in any coordinate system) is $\mu 1$. If $\alpha \to \bar{\alpha}$ and T are the automorphism and matrix of $\mathbf{T}$ relative to $(e_1, \cdots, e_n)$, $\mathbf{T}^m$ is the s.l.t. whose automorphism and matrix are

$$\alpha \to \bar{\alpha}^{(m)} \qquad \text{and} \qquad T\bar{T} \cdots \bar{T}^{(m-1)}.$$

Hence n.a.s.c.'s that an s.l.t. be periodic are

$$(34) \qquad \bar{\alpha}^{(m)} = \mu^{-1}\alpha\mu \qquad\qquad T\bar{T} \cdots \bar{T}^{(m-1)} = \mu 1.$$

Thus $\alpha \to \bar{\alpha}$ has relative finite order, say q and $m = qr$. (34) implies.

$$\bar{\mu}1 = \bar{T} \cdots \bar{T}^{(m)} = T^{-1}(T \cdots \bar{T}^{(m-1)})\bar{T}^{(m)} = T^{-1}\mu 1 \bar{T}^{(m)} = \mu 1$$

and so $\mu^*(t) = t^m - \mu \in \mathfrak{B}^*$. $\mathbf{T}$ is a finite s.l.t. whose minimum function is a factor of $\mu^*(t)$.

THEOREM 20. *If $\chi(\mathfrak{F}) = p$ is not a divisor of $m/q = r$, $\mathbf{T}$ is completely reducible.*

If $\bar{\alpha}^{(q)} = \nu\alpha\nu^{-1}$, $u = t^q\nu$ commutes with every α. $t^m - \mu = u^r\rho - \mu$ where $\rho^{-1} = \bar{\nu}^{(q(r-1))} \cdots \bar{\nu}^{(q)}\nu$. Since $\mu \neq 0$, the factors of $u^r\rho - \mu$ in $\mathfrak{B}^*$ are polynomials in u. Suppose

$$(35) \qquad u^r\rho - \mu = \pi_1^*(u)^2\pi_2^*(u) \qquad\qquad \pi_i^*(u) \in \mathfrak{B}^*.$$

Since u commutes with the coefficients of these polynomials, we may differentiate (35) formally with respect to u and obtain

$$ru^{r-1}\rho = [\pi_1^*(u)\pi_1^*(u)' + \pi_1^*(u)'\pi_1^*(u)]\pi_2^*(u) + \pi_1^*(u)^2\pi_2^*(u)' \equiv 0 \ (\pi_1^*(u))$$

since $\pi_1^*(u) \in \mathfrak{B}^*$. If $r \not\equiv 0 \pmod p$, this implies $(u^{r-1}, u^r\rho - \mu) \neq 1$ which is impossible since $\mu \neq 0$. Hence $t^m - \mu$ has no multiple factors in $\mathfrak{B}^*$ and $\mathbf{T}$ is completely reducible.

If $m = q$, $\mathbf{T}$ will be called *normal*. In the remainder of this section we suppose that $\mathbf{T}$ is a normal periodic s.l.t. In this case $\mu^*(t) = t^q - \mu$ is irreducible in $\mathfrak{B}^*$. Hence $\mu^*(t)$ is the minimum function of $\mathbf{T}$ which is then completely reducible. If

$$\mu(t) = t^l - t^{l-1}\mu_1 - \cdots - \mu_l$$

is an irreducible factor of $\mu^*(t)$ in $\mathfrak{F}[t]$, then $l \mid m$, $m = m'l$ and a matrix of $\mathbf{T}$ has the form

$$(36) \qquad \begin{pmatrix} U & & & \\ & U & & \\ & & \ddots & \\ & & & U \end{pmatrix}$$

where

$$\begin{pmatrix} 0 & \cdot & \cdot & \cdot & \cdot & \mu_l \\ 1 & & & & & \cdot \\ & \cdot & & & & \cdot \\ & & \cdot & & & \cdot \\ & & & \cdot & & \mu_2 \\ & & & & 1 & \mu_1 \end{pmatrix} \tag{37}$$

Thus $l \mid n$, $n = n'l$. We shall call l the *index* of **T**.

T is a multiple of the identity matrix if and only if $l = 1$ i.e. $\mu(t) = t - \mu_l \equiv t - \nu$. Then

$$t^q - \mu = (t - \nu)(t^{q-1} + t^{q-2}\bar{\nu}^{(q-1)} + \cdots + \bar{\nu}\bar{\nu}^{(2)} \cdots \bar{\nu}^{(q-1)})$$

and hence

$$\mu = N(\nu) = \nu\bar{\nu} \cdots \bar{\nu}^{(q-1)}$$

and conversely if $\mu = N(\nu)$, **T** has index $= 1$.

THEOREM 21. *A n.a.s.c. that a normal periodic s.l.t. have index* 1 *is that* $\mu = N(\nu)$ *for some* ν *in* $\mathfrak{F}$.

In particular if $\mu = 1$, we may take $\nu = 1$ and then (36) is the identity matrix. In the language of matrices this becomes the following extension of Hilbert's theorem on cyclic fields.[18]

COROLLARY. *Let* $\mathfrak{F}$ *be any field and* $\alpha \rightarrow \bar{\alpha}$ *any automorphism whose order and relative order are finite and* $= q$. *If* T *is a matrix with elements in* $\mathfrak{F}$ *such that* $T\bar{T} \cdots \bar{T}^{(q-1)} = 1$, *then there exists a matrix* A *in* $\mathfrak{F}$ *such that* $T = A^{-1}\bar{A}$.

The automorphism ring $\mathfrak{A}$ of **T** with matrix (36) is according to the general theory a simple algebra $\mathfrak{D}_{n'}$ where $\mathfrak{D}$ is inverse isomorphic to the invariant ring of $\mu(t)$. If **T** has the index 1 ($\mu(t) = t - \nu$). It follows from the definition of the invariant ring that $\mathfrak{D}$ is inverse isomorphic to the subfield of $\mathfrak{F}$ of elements β such that $\bar{\beta} = \nu^{-1}\beta\nu$.

If $n = q$, **T** is cyclic and $\mathfrak{A}$ is inverse isomorphic to the invariant ring of $\mu^*(t) = t^q - \mu$. The latter is a generalization of cyclic algebras[19] first considered by Dickson. The above results are applicable to the study of its structure. Thus we have the following extension of a theorem conjectured by Dickson and proved by Albert.[20]

THEOREM 22. *If* q *is a prime, the invariant ring* $\mathfrak{A}$ *of* $(t^q - \mu) \epsilon \mathfrak{B}^*$ *is a field if* $\mu \neq N(\nu)$ *for any* $\nu \epsilon \mathfrak{F}$. *If* $\mu = N(\nu)$, $\mathfrak{A} \simeq \mathfrak{D}_q$ *where* $\mathfrak{D}$ *is the subfield of* β*'s such that* $\bar{\beta} = \nu^{-1}\beta\nu$.

12. **Differential transformations.** In this section we suppose that $\mathfrak{F}$ is commutative and **T** is a d.t. in $\mathfrak{R}$ over $\mathfrak{F}$. We denote the subfield of *constants* (elements β such that $\beta' = 0$) of $\mathfrak{F}$ by $\mathfrak{K}$.

---

[18] Hilbert [5] p. 272.

[19] A discussion of cyclic algebras from the point of view of the invariant ring of non-commutative polynomials is given in Jacobson [6].

[20] Dickson [3] p. 227 and Albert [1] p. 624. Our extension consists in removing the assumption that $\mathfrak{F}$ has a finite basis over its centrum.

We have seen that the automorphism ring $\mathfrak{A}$ of $\mathbf{T}$ is given by the matrices A such that

$$[A, T] \equiv AT - TA = A' \tag{38}$$

where T is a matrix of $\mathbf{T}$. This matrix equation is equivalent to a system of n^2 linear homogeneous differential equations in the elements of A. Hence the number of solutions of (38) which are linearly independent relative to $\mathfrak{K}$ is finite. We therefore have the following theorem due to Ore[21] for cyclic d.t.

THEOREM 23. *If* $\mathbf{T}$ *is a d.t. in* $\mathfrak{R}$ *over a commutative field* $\mathfrak{F}$, *its automorphism ring has a finite basis over the field of constants* $\mathfrak{K}$.

Suppose $\mathfrak{K}$ is perfect and let $\mathfrak{K}(\theta)$ be an algebraic extension of $\mathfrak{K}$ of degree r. Consider $\mathfrak{F}(\theta)$. Since θ is a separable element, we may extend the derivative defined in $\mathfrak{F}$ to $\mathfrak{F}(\theta)$ by defining $\theta' = 0$. It follows that $\mathfrak{F}(\theta)$ has degree r over $\mathfrak{F}$ also; for if $\theta^m + \alpha_1\theta^{m-1} + \cdots + \alpha_m = 0$ is the minimum equation of θ over $\mathfrak{F}$, we obtain by differentiation $\alpha_1'\theta^{m-1} + \cdots + \alpha_m' = 0$. Hence each $\alpha_i' = 0$ or $\alpha_i \in \mathfrak{K}$ and $m = r$.

Let $\mathfrak{R}_{\mathfrak{F}(\theta)}$ be the extension of $\mathfrak{R}$ obtained by letting the coefficients vary in $\mathfrak{F}(\theta)$. If $\mathbf{T}$ is a d.t. in $\mathfrak{R}$, it has a unique extension $\mathbf{T}(\theta)$ in $\mathfrak{R}_{\mathfrak{F}(\theta)}$: The matrix T of $\mathbf{T}$ relative to $(e_1, \cdots, e_n)$ is also the matrix of $\mathbf{T}(\theta)$ relative to this basis. It is easily seen that *the automorphism ring of* $\mathbf{T}(\theta)$ *is* $\mathfrak{A}_{\mathfrak{F}(\theta)}$ *where* $\mathfrak{A}$ *is that of* $\mathbf{T}$; for an automorphism of $\mathbf{T}(\theta)$ is determined by $A(\theta)$ such that

$$[A(\theta), T] = A(\theta)'. \tag{39}$$

If $\theta_1, \cdots, \theta_r$ is a basis for $\mathfrak{F}(\theta)$ over $\mathfrak{F}$, we may write $A(\theta)$ uniquely in the form

$$A_1\theta_1 + A_2\theta_2 + \cdots + A_r\theta_r \qquad A_i \in \mathfrak{F}_n.$$

Since $\theta_i' = 0$, (39) gives $[A_i, T] = A_i'$, i.e. A_i determine automorphism of $\mathbf{T}$. The converse of this is obvious.

Now suppose $\mathbf{T}$ is an irreducible d.t. which decomposes in every algebraic extension $\mathfrak{F}(\theta)$ of $\mathfrak{F}$ into parts all of which are similar. Then the automorphism ring $\mathfrak{A}$ of $\mathbf{T}$ is a normal division algebra over $\mathfrak{K}$. For if the centrum of $\mathfrak{A}$ has degree m over $\mathfrak{K}$, it follows from the theory of hypercomplex numbers that there exists a field $\mathfrak{K}(\theta)$ such that $\mathfrak{A}_{\mathfrak{F}(\theta)}$ is a direct sum of m matrix algebras. But then $\mathbf{T}(\theta)$ would have dissimilar components unless $m = 1$. This criterion may be used to construct examples of d.t.'s whose automorphism rings are normal division algebras.

BIBLIOGRAPHY

1. A. A. ALBERT, *On normal simple algebras*, Trans. Am. Math. Soc., **34** (1932), pp. 620–625.
2. G. BIRKHOFF, *On the combination of subalgebras*, Proc. Camb. Phil. Soc., **29** (1933), pp. 441–464.
3. L. E. DICKSON, *New division algebras*, Trans. Am. Math. Soc., **28** (1926), pp. 207–234.
4. H. FITTING, *Die Theorie der Automorphismenringe Abelscher Gruppen* etc., Math. Annalen, **107** (1932), pp. 514–542.

[21] Ore [11] p. 236.

5. D. Hilbert, *Die Theorie der algebraische Zahlkorper*, Jahr. Deutsch. Math. Verein., 4 (1897).
6. N. Jacobson, *Non-commutative polynomials and cyclic algebras*, Annals of Math. **35** (1934), pp. 197–209.
7. N. Jacobson, *On pseudo-linear transformations*, Proc. Nat. Acad. Sci., **21** (1935), pp. 667–670.
8. W. Krull, *Über verallgemeinerte endliche Abelsche Gruppen*, Math. Zeitsch., **23** (1925), pp. 160–196.
9. W. Krull, *Theorie und Anwendung der verallgemeinerten Abelschen Gruppen*, Sitz. Heidelberg Akad., **17** (1926), pp. 3–32.
10. E. Noether and W. Schmeidler, *Moduln in nichtkommutative Bereichen* etc., Math. Zeitsch., **8** (1920), pp. 1–35.
11. O. Ore, *Formale theorie lineare differentialgleichungen* II, Jour. für Math., **168** (1932), pp. 233–252.
12. O. Ore, *Theory of non-commutative polynomials*, Annals of Math., **34** (1933), pp. 480–508.
13. O. Ore, *On the foundations of abstract algebra* I, Annals of Math., **36** (1935), pp. 406–437.
14. C. Segre, *Un nuovo campo di ricerche geometriche*, Atti Torino **25** (1889), pp. 276–301.
15. B. L. van der Waerden, *Moderne Algebra* I and II, J. Springer, Berlin, 1930.
16. B. L. van der Waerden, *Gruppen von lineare Transformationen*, Ergebnisse der Math., J. Springer, Berlin, 1935.
17. J. H. M. Wedderburn, *Non-commutative domains of integrity*, Jour. für Math., **167** (1932), pp. 129–141.
18. J. H. M. Wedderburn, *Lectures on matrices*, Am. Math. Soc. Colloquium Publications, vol. XVII, New York, 1934.

Bryn Mawr College.

Annals of Mathematics
Vol. 38, No. 2, April, 1937

A CLASS OF NORMAL SIMPLE LIE ALGEBRAS OF CHARACTERISTIC ZERO[1]

By N. Jacobson[2]

(Received March 24, 1937)

Introduction. The classification of simple Lie algebras (infinitesimal groups) over the field of complex numbers is due to Killing and to Cartan.[3] In a second paper[4] Cartan determined the simple Lie algebras over the real field. The methods used in these papers are purely algebraic and so it is clear that the results hold equally well for arbitrary algebraically closed fields of characteristic zero and arbitrary real closed fields. More recently W. Landherr[5] considered this problem for arbitrary fields of characteristic zero and showed that it could be reduced to that of determining all *normal simple algebras* i.e. algebras which remain simple when the underlying field is extended to its algebraic closure. It therefore suffices to classify those Lie algebras which become a fixed Lie algebra in the Killing-Cartan list when the field is extended. Landherr has obtained a partial solution of this problem for the algebras which lead to Cartan's type A.[6] In this paper we determine the algebras which lead to Cartan's types B, C, D.[7] We give an invariantive realization of these systems and also obtain their groups of automorphisms. The relation between the Lie algebra and this group seems to be an algebraic substitute for the relation between an infinitesimal group and the integrated group given by Lie's theory.

1. Anti-automorphisms in an algebra.

Let $\mathfrak{A}$ be an associative algebra with a finite basis over a commutative field Φ. It is readily verified that if we define in $\mathfrak{A}$ the new composition $[a, b] = ab - ba$, $\mathfrak{A}$ becomes a Lie algebra i.e.

$$\alpha[a, b] = [\alpha a, b] = [a, \alpha b] \qquad [a + b, c] = [a, c] + [b, c]$$

$$[a, b] = -[b, a] \qquad [[a, b], c] + [[b, c], a] + [[c, a], b] = 0.$$

Suppose $\mathfrak{A}$ is *self-reciprocal* i.e. there is defined a correspondence (an *anti-automorphism*) J, $a \to a^J$ in $\mathfrak{A}$ such that

$$(a + b)^J = a^J + b^J \qquad (\alpha a)^J = \alpha a^J \qquad (ab)^J = b^J a^J.$$

---

[1] Presented to the Society, Dec. 29, 1936.
[2] National Research Fellow.
[3] E. Cartan, *Thèse*, Paris (1894) Chapter 5.
[4] *Les groupes réels simples et continus*, Annales de l'École Normale, **31** (1914) pp. 263–355.
[5] *Über einfache Liesche Ringe*, Hamb. Abhandlungen **11** (1935) pp. 41–64.
[6] His classification is complete for p-adic fields.
[7] With the exception of type D and order 28.

We shall call a *J-skew* if $a^J = -a$. Let $\mathfrak{S}_J$ denote the set of J-skew elements. If $a, b \in \mathfrak{S}_J$ so do $a + b$, αa and $[a, b]$ so that $\mathfrak{S}_J$ is a Lie subalgebra of $\mathfrak{A}$.

A second anti-automorphism K is *cogredient* to J if $K = S^{-1}JS$ where S is an automorphism of $\mathfrak{A}$. In this case the elements of $\mathfrak{S}_K$ are a^S, a in $\mathfrak{S}_J$ ($\mathfrak{S}_K = \mathfrak{S}_J^S$) and the correspondence $a \rightarrow a^S$ establishes an isomorphism between the Lie algebras $\mathfrak{S}_J$ and $\mathfrak{S}_K$.

The elements g of $\mathfrak{A}$ such that $gg^J = \gamma 1 = g^J g$ where $\gamma \neq 0 \in \Phi$ will be called *J-orthogonal.* Their totality is a group $\mathfrak{G}_J$ under multiplication. $\mathfrak{G}_J$ contains the invariant subgroup $\mathfrak{D}$ consisting of the Φ-multiples of 1. We note that $g^{J^2} = (\gamma g^{-1})^J = \gamma(g^J)^{-1} = \gamma\gamma^{-1}g = g$ and $(g^{-1}ag)^J = g^J a^J (g^J)^{-1} = g^{-1}a^J g$. Hence if a is J-skew so is $g^{-1}ag$ and so if g is a fixed element of $\mathfrak{G}_J$ the correspondence $a \rightarrow g^{-1}ag$ is an automorphism of the Lie algebra $\mathfrak{S}_J$. If S is an automorphism of $\mathfrak{A}$ then $\gamma 1 = g^S g^{JS} = g^S g^{S(S^{-1}JS)} = g^S g^{SK}$, $K = S^{-1}JS$. Thus the correspondence $g \rightarrow g^S$ is an isomorphism between $\mathfrak{G}_J$ and $\mathfrak{G}_K$.

Let $\tilde{\Phi}$ be an over-field of Φ and $\xi_1, \xi_2, \cdots$ be a basis of $\tilde{\Phi}$ over Φ. The elements of the algebra $\tilde{\mathfrak{A}} = \mathfrak{A}_{\tilde{\Phi}}$ obtained from $\mathfrak{A}$ by extending Φ to $\tilde{\Phi}$ are uniquely expressible in the form $a = \xi_1 a_1 + \xi_2 a_2 + \cdots$ (finite sums) where $a_i \in \mathfrak{A}$. If M is an automorphism or anti-automorphism of $\mathfrak{A}$, M has a unique extension defined by $a^M = \xi_1 a_1^M + \xi_2 a_2^M + \cdots$ in $\tilde{\mathfrak{A}}$. If $M = J$ is an anti-automorphism and a is J-skew,

$$\xi_1 a_1^J + \xi_2 a_2^J + \cdots = -(\xi_1 a_1 + \xi_2 a_2 + \cdots)$$

then $a_i^J = -a_i$. Thus the J-skew elements of $\tilde{\mathfrak{A}}$ are the linear combinations with coefficients in $\tilde{\Phi}$ of the J-skew elements of $\mathfrak{A}$ i.e. $\mathfrak{S}_{\tilde{\Phi}J} = \mathfrak{S}_{J\tilde{\Phi}}$.

J is *involutorial* if $J^2 = I$ the identity mapping. If K is cogredient to J then K is involutorial also.

2. **Skew elements of a matric algebra.** In the remainder of the paper J, K will denote involutorial anti-automorphism (i. a. a.), $\mathfrak{A}$ a normal simple algebra over Φ of characteristic 0.

Suppose first that $\mathfrak{A} \cong \Phi_n$ the matrix algebra of order n^2 over Φ. The correspondence J_0: $A = (\alpha_{ij}) \rightarrow A' = (\alpha_{ji})$ is an i. a. a.[8] If J is any anti-automorphism $J = SJ_0$ where S is an automorphism and since any automorphism of a normal simple algebra is inner[9] we have $A^J = S^{-1}A'S$ where $A \rightarrow S^{-1}AS$ is the automorphism S. The condition $J^2 = I$ is $S^{-1}S'A(S^{-1}S') = A$ for all A. Hence $S^{-1}S' = \sigma 1$ and $S' = \sigma S$ and since $S'' = S$, $\sigma^2 = 1$ and $\sigma = \pm 1$. Thus S is either symmetric or skew-symmetric (relative to J_0).[10] The i. a. a. $K = V^{-1}JV$ is given as $A^K = T^{-1}A'T$ where $T = V'SV$ and $A^V =$

---

[8] J_0 is defined by means of a fixed basis E_{ij}, of Φ_n such that $E_{ij}E_{kl} = \delta_{jk}E_{il}$: we have $E_{ij}^{J_0} = E_{ji}$.

[9] For a proof of this theorem see M. Deuring, *Algebren*, Springer 1935, p. 42.

[10] Cf. A. A. Albert, *Involutorial simple algebras and real Riemann matrices*, Annals of Math. **36** (1935) p. 897.

$V^{-1}AV$ as is readily verified. It should be noted that the matrix S associated with J is determined only to within a multiple in Φ. Hence we have

THEOREM 1. *The i. a. a. J and K are cogredient if and only if the matrix S of J is cogredient (in the usual sense) to a multiple of the matrix T of K.*

(1) If $S' = -S$, since S is non-singular $n = 2\nu$, ν an integer and S is cogredient to

$$\begin{pmatrix} 0 & 1_\nu \\ -1_\nu & 0 \end{pmatrix}$$

(1_ν the identity matrix of Φ_ν).[11] The corresponding i. a. a. K defines the Lie algebra $\mathfrak{S}_K$ consisting of the matrices of the form

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

where $A_{ij} \in \Phi_\nu$ and $A_{11}' = -A_{22}$, $A_{12}' = A_{12}$, $A_{21}' = A_{21}$.[12] Thus $\mathfrak{S}_K$ has order $n(n+1)/2$ and is generated by the abelian subalgebra $\mathfrak{H}$ of matrices

$$H = \begin{pmatrix} \lambda_1 & & & & & \\ & \ddots & & & & \\ & & \lambda_\nu & & & \\ & & & -\lambda_1 & & \\ & & & & \ddots & \\ & & & & & -\lambda_\nu \end{pmatrix} \tag{1}$$

and

$$E_{\lambda_i-\lambda_k} = \begin{pmatrix} E_{ik} & 0 \\ 0 & -E_{ki} \end{pmatrix} \qquad E_{-\lambda_i+\lambda_k} = \begin{pmatrix} E_{ki} & 0 \\ 0 & -E_{ik} \end{pmatrix}$$

$$E_{\lambda_i+\lambda_k} = \begin{pmatrix} 0 & E_{ik}+E_{ki} \\ 0 & 0 \end{pmatrix} \qquad E_{-\lambda_i-\lambda_k} = \begin{pmatrix} 0 & 0 \\ E_{ik}+E_{ki} & 0 \end{pmatrix}$$

$$E_{2\lambda_i} = \begin{pmatrix} 0 & E_{ii} \\ 0 & 0 \end{pmatrix} \qquad E_{-2\lambda_i} = \begin{pmatrix} 0 & 0 \\ E_{ii} & 0 \end{pmatrix}$$

where $i < k$ and E_{ij} is the matrix in Φ_ν having a 1 in the i^{th} row and j^{th} column. The linear forms $\alpha = \pm\lambda_i \pm \lambda_k$, $\pm 2\lambda_i$ are the *roots* of $\mathfrak{S}_K$. The multiplication table of this algebra is

---

[11] See, for example J. H. M. Wedderburn, *Lectures on matrices*, New York, 1934, p. 96.

[12] $\mathfrak{S}_K$ is the "infinitesimal group" of the complex group when Φ is the field of complex numbers. Cf. H. Weyl, *Theorie der Darstellung kontinuierlicher halbeinfacher Gruppen* II, Math. Zeitschrift **24** (1926) p. 333.

$$
(2)\qquad
\begin{aligned}
&[H_1, H_2] = 0 \text{ for any } H_1, H_2 \text{ in } \mathfrak{H}\\
&[H, E_\alpha] = \alpha E_\alpha \qquad [E_\alpha, E_{-\alpha}] = H_\alpha \in \mathfrak{H}\\
&[E_\alpha, E_\beta] = \begin{cases} 0 \text{ if } \alpha + \beta \text{ is not a root}\\ N_{\alpha\beta}E_{\alpha+\beta} \neq 0 \;\; N_{\alpha\beta} \text{ an integer if } \alpha + \beta \text{ is a root.}\end{cases}
\end{aligned}
$$

It follows that $\mathfrak{S}_K$ *is a simple Lie algebra*. For if $\mathfrak{J} \neq 0$ is an ideal in $\mathfrak{S}_K$ and $B = H_0 + \sum \kappa_\alpha E_\alpha \neq 0$ is in $\mathfrak{J}$ then so is

$$
(3)\qquad \underbrace{[H, [H, \cdots [H}_{r}, B] \cdots]] = \sum \kappa_\alpha \alpha^r E_\alpha \qquad r = 1, 2, \cdots
$$

for all values of $\lambda_1, \cdots, \lambda_\nu$. The determinant of $m = 2\nu^2$ rows (m, the number of roots)

$$
V(\lambda_1, \cdots, \lambda_\nu) = \begin{vmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_m\\ \alpha_1^2 & \alpha_2^2 & \cdots & \alpha_m^2\\ \cdot & \cdot & & \cdot\\ \cdot & \cdot & & \cdot\\ \cdot & \cdot & & \cdot\\ \alpha_1^m & \alpha_2^m & \cdots & \alpha_m^m \end{vmatrix} = \prod_j \alpha_j \prod_{i<k} (\alpha_i - \alpha_k)
$$

where the α_i are the roots is not identically 0 and so $\lambda_1^0, \cdots, \lambda_\nu^0$ can be chosen in Φ so that $V(\lambda_1^0, \cdots, \lambda_\nu^0) \neq 0$. By multiplying the relations (3) by the cofactors of the column containing α in V and adding we obtain that

$$
V(\lambda_1^0, \cdots, \lambda_\nu^0)\kappa_\alpha E_\alpha \in \mathfrak{J}
$$

and hence also $\kappa_\alpha E_\alpha$. Thus if $\kappa_\alpha \neq 0$, $E_\alpha \in \mathfrak{J}$. If $\nu > 1$ every root may be obtained from a fixed one by adding roots and hence by the third relation (2) all E_α belong to $\mathfrak{J}$. It follows that all $H_\alpha \in \mathfrak{J}$ and since the H_α generate $\mathfrak{H}$, $\mathfrak{J} = \mathfrak{S}_K$. If $\nu = 1$, $\mathfrak{J} \supset [E_\alpha, E_{-\alpha}]$ which forms a basis for $\mathfrak{H}$ and hence $\mathfrak{J} \supset E_{-\alpha}$ also. If all $\kappa_\alpha = 0$, $H_0 \neq 0$ and some $[H_0, E_\alpha]$ is a non-zero multiple of E_α. As before it follows that $\mathfrak{J} = \mathfrak{S}_K$.

(2) $S' = S$. In this case S is cogredient to

$$
(4)\qquad \begin{pmatrix} \rho_1 & & & \\ & \rho_2 & & \\ & & \ddots & \\ & & & \rho_n \end{pmatrix}
$$

and if Φ contains $\sqrt{\rho_i}$, S is cogredient to 1_n and also to either

$$
(5)\qquad \begin{pmatrix} 1 & & \\ & 0 & 1_\nu\\ & 1_\nu & 0 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 0 & 1_\nu\\ 1_\nu & 0 \end{pmatrix}
$$

according as $n = 2\nu + 1$ or $n = 2\nu$.[13] If we use the i. a. a. corresponding to 1_n we see that $\mathfrak{S}_J$ is isomorphic to the Lie algebra of all skew-symmetric matrices in Φ_n and so $\mathfrak{S}_J$ has order $n(n - 1)/2$. For our purposes it is more convenient to take the i. a. a. defined by the matrices (5) and to distinguish the two cases odd and even order. A discussion analogous to that given in (1) may be made. We shall merely state the results referring the reader to Prof. Weyl's paper[14] for the details.

(a) $n = 2\nu + 1$. $\mathfrak{H}$ is the subalgebra of matrices

$$H = \begin{pmatrix} 0 & & & & & & \\ & \lambda_1 & & & & & \\ & & \ddots & & & & \\ & & & \lambda_\nu & & & \\ & & & & -\lambda_1 & & \\ & & & & & \ddots & \\ & & & & & & -\lambda_\nu \end{pmatrix}$$

and there are E_α's such that (2) holds. The roots α are $\pm\lambda_i \pm\lambda_k$ $(i < k)$ and $\pm\lambda_i$. As before $\mathfrak{S}_K$ is simple.

(b) $n = 2\nu$. $\mathfrak{H}$ is the subalgebra

$$H = \begin{pmatrix} \lambda_1 & & & & & \\ & \ddots & & & & \\ & & \lambda_\nu & & & \\ & & & -\lambda_1 & & \\ & & & & \ddots & \\ & & & & & -\lambda_\nu \end{pmatrix}$$

and the roots α corresponding to the E_α are $\pm\lambda_i \pm\lambda_k$ $(i < k)$. The proof given above for the simplicity of $\mathfrak{S}_K$ in (1) is valid here for $\nu > 2$. If $\nu = 1$ $\mathfrak{S}_K$ is abelian and if $\nu = 2$ $\mathfrak{S}_K$ is a direct sum of two algebras of type $2a$ in which $\nu = 1$.

It can be shown that $\mathfrak{S}_K$ of type 2a with $\nu = 1$ is isomorphic to $\mathfrak{S}_K$ of type 1 for $\nu = 1$ and type 2a with $\nu = 2$ is isomorphic to type 1 with $\nu = 2$. No other algebra of type 2 is isomorphic with one of type 1.[15]

In either (1) or (2) the set of matrices $\mathfrak{S}_K$ are an absolutely irreducible system.[16] By Burnside's theorem the enveloping algebra of $\mathfrak{S}_K$ (the smallest subalgebra of Φ_n containing $\mathfrak{S}_K$) is Φ_n.

---

[13] Wedderburn, *Lectures on matrices*, p. 95.

[14] Loc. cit. in 12, pp. 342–345.

[15] Cartan, *Thèse*, p. 78.

[16] Weyl II, p. 334 and p. 344.

3. **Skew elements of a normal simple algebra.** Now suppose that $\mathfrak{A}$ is a normal simple algebra of degree n (order n^2) and Ω is the algebraic closure of Φ. Then $\mathfrak{A}_\Omega \cong \Omega_n$ and $\mathfrak{A}$ has a representation $a \to A$ by matrices in Ω such that the linear combinations $\sum \tilde{\alpha}_i A_i$ with coefficients $\tilde{\alpha}_i$ in Ω of the matrices (representing the elements) of $\mathfrak{A}$ comprise all the elements of Ω_n. The i. a. a. J in $\mathfrak{A}$ induces an i. a. a. J in Ω_n $A \to A^J$ where A^J corresponds to a^J, $\sum \tilde{\alpha}_i A_i \to \sum \tilde{\alpha}_i A_i^J$. The J-skew elements of Ω_n are the linear combinations with coefficients in Ω of the matrices of $\mathfrak{S}_J$. Now let G be a matrix of $\mathfrak{A}$ such that G is J-orthogonal in Ω_n i.e. $GG^J = \tilde{\gamma}1$, $\tilde{\gamma} \neq 0$ in Ω. Then GG^J is a matrix of $\mathfrak{A}$ and commutes with all the matrices of $\mathfrak{A}$. Hence $\tilde{\gamma} = \gamma \in \Phi$ and G is the matrix of a J-orthogonal element of $\mathfrak{A}$.

We have seen that there is an automorphism S in Ω_n such that $K = S^{-1}JS$ is one of the normalized i. a. a.'s defined in the last section and $\tilde{\mathfrak{S}}_K$ the set of K-skew elements of Ω_n is one of the systems (1), (2a) or (2b). The correspondence $a \to A^S$ is a second representation of $\mathfrak{A}$ by matrices in Ω_n and the i. a. a. J in $\mathfrak{A}$ induces by means of this representation the i. a. a. K in Ω_n since $A^S \to A^{JS} = A^{SK}$. We may restrict our attention to the second representation and write $a \to A$ in place of $a \to A^S$ and J in place of K. We shall say that $\mathfrak{S}_J$ has type (1), (2a) or (2b) according as $\tilde{\mathfrak{S}}_J$ is the system (1), (2a) or (2b). If $\mathfrak{S}_J$ has type (1) its order is $n(n + 1)/2$ and if it has types (2) its order is $n(n - 1)/2$. In either case since the enveloping algebras (1), (2a) or (2b) is Ω_n, the enveloping algebra of $\mathfrak{S}_J$ is $\mathfrak{A}$.

A Lie algebra $\mathfrak{L}$ over Φ will be called *normal simple* if $\mathfrak{L}_\Omega$ where Ω is the algebraic closure of Φ is simple. This of course implies that $\mathfrak{L}$ is simple. For if $\mathfrak{I}$ is an ideal in $\mathfrak{L}$, $\mathfrak{I}_\Omega$ is one in $\mathfrak{L}_\Omega$. We have seen that all the algebras $\mathfrak{S}_J$ with the exception of type 2 for $n = 4$ are normal simple. We summarize our results in the theorem:

THEOREM 2. *If $\mathfrak{A}$ is a normal simple algebra of degree n and J an i. a. a. in $\mathfrak{A}$, $\mathfrak{S}_J$ the set of J-skew elements is a normal simple algebra except when it has type 2 and $n = 4$. The order of $\mathfrak{S}_J$ is $n(n + 1)/2$ (type 1) or $n(n - 1)/2$ (type 2). In either case the enveloping algebra of $\mathfrak{S}_J$ is $\mathfrak{A}$.*

COROLLARY. *If J and K are i. a. a. in a normal simple algebra and $\mathfrak{S}_J = \mathfrak{S}_K$ then $J = K$.*

For the elements a of $\mathfrak{S} = \mathfrak{S}_J = \mathfrak{S}_K$ we have $a^J = -a = a^K$. Since the enveloping algebra of $\mathfrak{S}$ is $\mathfrak{A}$ any element b in $\mathfrak{A}$ has the form $\sum a_1 a_2 \cdots a_r$, $a_i \in \mathfrak{S}$. Hence $b^J = \sum (-1)^r a_r a_{r-1} \cdots a_1 = b^K$, i.e. $J = K$.

4. **Automorphisms of the Lie algebra $\mathfrak{S}_J$.** From now on we suppose when $\mathfrak{S}_J$ has type 2 that $n \geqq 6$ and $\neq 8$. In view of the isomorphisms noted in §2 the only loss in generality here is in the restriction $n \neq 8$.

Let $G: a \to a^G$ be an automorphism $\neq 0$ of $\mathfrak{S}_J$. G is $(1 - 1)$ since $\mathfrak{S}_J$ is simple and the elements mapped into 0 by an automorphism form an ideal. We shall show that G may be realized as a similarity transformation by a J-orthogonal element. For this purpose we suppose first that $\Phi = \Omega$ is algebraically closed. Then $\mathfrak{A} \cong \Omega_n$ and $\mathfrak{S}_J$ may be taken to be one of the systems (1), (2a) or (2b).

If A is a matrix of $\mathfrak{S}_J$ the correspondence $A \rightarrow A^G$ is a second representation of $\mathfrak{S}_J$ and since the totality of matrices $\{A^G\}$ coincides with the totality $\{A\}$ this representation is irreducible also. If, as above, $H = H(\lambda_1, \cdots, \lambda_\nu)$ denotes the general element of the maximal abelian subalgebra $\mathfrak{H}$, by Cartan's theory of representations there exists a matrix Q in Ω_n such that

$$K = Q^{-1}H^G Q = \begin{pmatrix} \Lambda_1 & & \\ & \Lambda_2 & \\ & & \ddots \end{pmatrix}$$

where the Λ's the weights of the representation are linear forms in the λ's with coefficients all integers in (1) and either all integers or all halves of odd integers in (2).[17] Since every A of $\mathfrak{S}_J$ is similar to $-A'$, if Λ_0 is a root of the characteristic equation of A so is $-\Lambda_0$. It follows that the negative of every weight is a weight. For if no $\Lambda_k = -\Lambda_p$, λ_i may be chosen $= \lambda_i^0$ in Ω so that

$$\Lambda_p(\lambda_1^0, \cdots, \lambda_\nu^0) \neq 0, \neq -\Lambda_k(\lambda_1^0, \cdots, \lambda_\nu^0).$$

Suppose $Q^{-1}E_\alpha^G Q = F_\alpha = (\varphi_{ij\alpha})$ where E_α is defined as in §2. Since $[H, E_\alpha] = \alpha E_\alpha$, $[K, F_\alpha] = \alpha F_\alpha$ and so $\alpha\varphi_{ij\alpha} = (\Lambda_i - \Lambda_j)\varphi_{ij\alpha}$. Since $F_\alpha \neq 0$ it follows that every root α is expressible as a difference of weights. Thus the number of linearly independent weights is ν. We order them lexicographically[18] and may suppose that $\Lambda_1 > \Lambda_2 > \cdots > \Lambda_\nu > 0 > -\Lambda_\nu > \cdots > -\Lambda_1$. We recall also that the set of weights is invariant under arbitrary permutations of the λ's.[19]

(1). In this case the weights are $\pm\Lambda_1, \cdots, \pm\Lambda_\nu$ and the possible differences $\neq 0$ are $\pm\Lambda_p \pm \Lambda_q$, $p \neq q$ and $\pm 2\Lambda_r$. Since their number is equal to the number of roots all of these differences are roots. Evidently $2\Lambda_1 > \pm\Lambda_p \pm \Lambda_q > \pm 2\Lambda_r$ if $r \neq 1$ and so $2\Lambda_1 = 2\lambda_1$ the highest root and $\Lambda_1 = \lambda_1$.

(2a). The weights here are $0, \pm\Lambda_1, \cdots, \pm\Lambda_\nu$ and the possible differences are $\pm\Lambda_p \pm \Lambda_q$, $\pm\Lambda_r$, $\pm 2\Lambda_r$. If $2\Lambda_r$ is a root $\Lambda_r = \lambda_i/2$ or $= (\lambda_i \pm \lambda_k)/2$ $(i < k)$ and since the coefficients must all be halves of odd integers, $\nu = 1$ or 2 which are cases outside of the present consideration. Thus no $\pm 2\Lambda_r$ is a root and so all $\pm\Lambda_p \pm \Lambda_q$, $\pm\Lambda_r$ are. Comparing highest linear forms we have $\Lambda_1 + \Lambda_2 = \lambda_1 + \lambda_2$. If $\Lambda_1 - \Lambda_2 = \lambda_i$, $\Lambda_1 = (\lambda_1 + \lambda_2 + \lambda_i)/2$ and in order that the coefficients all be halves of odd integers $i \neq 1, 2$ and so $\nu = 3$, $i = 3$. Thus $\Lambda_2 = (\lambda_1 + \lambda_2 - \lambda_3)/2$ and by permuting the λ's and changing all signs we see that $(\lambda_1 - \lambda_2 + \lambda_3)/2$ and $(\lambda_1 - \lambda_2 - \lambda_3)/2$ are also weights. But this is impossible since there are only $\nu = 3$ positive weights. It follows that $\Lambda_1 - \Lambda_2 = \lambda_i \pm \lambda_k$ and $i < k$ since $\Lambda_1 - \Lambda_2$ is positive. If $k \neq 2$, $i \neq 1$ for otherwise the coefficients of $\Lambda_1 = (\lambda_1 + \lambda_2 + \lambda_i \pm \lambda_k)/2$ would not satisfy our arithmetic condition. Hence $\nu = 4$, $\Lambda_1 = (\lambda_1 + \lambda_2 + \lambda_3 \pm \lambda_4)/2$ and

[17] Cf. Weyl, *Darstellung kontinuierlicher halb-einfacher Gruppen* I, Math. Zeitschr. **23** (1925) p. 278 and Weyl II p. 334 and p. 344.

[18] We say that $\Lambda_i > \Lambda_j$ or $\Lambda_i - \Lambda_j = m_1\lambda_1 + \cdots + m_\nu\lambda_\nu > 0$ if the first $m_k \neq 0$ is positive.

[19] Weyl II, p. 334 and p. 344.

$\Lambda_2 = (\lambda_1 + \lambda_2 - \lambda_3 \mp \lambda_4)/2$. Then $\Lambda_3 = (\lambda_1 - \lambda_2 + \lambda_3 \mp \lambda_4)/2$ and $\Lambda_4 = (\lambda_1 - \lambda_2 - \lambda_3 \pm \lambda_4)/2$ are the remaining positive weights. But then $\pm\Lambda_p \pm \Lambda_q$, $\pm\Lambda_r$ do not exhaust all the roots and so this case is ruled out also. Hence $k = 2$, $i = 1$ and $\Lambda_1 - \Lambda_2 = \lambda_1 \pm \lambda_2$. Since $\Lambda_1 \neq 0$ $\Lambda_1 - \Lambda_2 = \lambda_1 - \lambda_2$ and $\Lambda_1 = \lambda_1$.

(2b). The weights are $\pm\Lambda_1, \cdots, \pm\Lambda_\nu$ and the possible differences are $\pm\Lambda_p \pm \Lambda_q$, $\pm 2\Lambda_r$. As in (2a) we may rule out the possibility that $\pm 2\Lambda_r$ are roots and so all $\pm\Lambda_p \pm \Lambda_q$ are. Again $\Lambda_1 + \Lambda_2 = \lambda_1 + \lambda_2$. As above $\Lambda_1 - \Lambda_2 = \lambda_i$ is impossible and $\Lambda_1 - \Lambda_2 = \lambda_i \pm \lambda_k$ $i \neq 1$, $k \neq 2$ gives $\nu = 4$ a case excluded by our assumption.[20] Hence $\Lambda_1 - \Lambda_2 = \lambda_1 - \lambda_2$ and again $\Lambda_1 = \lambda_1$.

Since the highest weight $\Lambda_1 = \lambda_1$ the representation by A^G is similar to that by A.[21] Thus there is a fixed matrix G in Ω_n such that $A^G = G^{-1}AG$ for all A. Since A^G and A are J-skew we have also $A^G = G^J A (G^J)^{-1}$ and hence GG^J commutes with all A. Since the system $\{A\}$ is irreducible $GG^J = \gamma 1$, $\gamma \neq 0$.

We may now prove the following theorem.

THEOREM 3. *If $\mathfrak{A}$ is a normal simple algebra, J an i. a. a. in $\mathfrak{A}$ and G an automorphism of the Lie algebra $\mathfrak{S}_J$ then there is a J-orthogonal element g such that $a^G = g^{-1}ag$ for all a.*[22]

We have seen in §3 that $\mathfrak{A}$ may be represented by matrices in Ω_n where Ω is the algebraic closure of Φ the field of $\mathfrak{A}$, in such a fashion that J induces in Ω_n a normalized i. a. a. of §2 and the linear combinations with coefficients in Ω of the matrices of $\mathfrak{S}_J$ is the system (*) = (1), (2a) or (2b). The automorphism G extends to an automorphism $\tilde{A} \to \tilde{A}^G$ of (*). We have seen that $\tilde{A}^G = \tilde{G}^{-1}\tilde{A}\tilde{G}$ where $\tilde{G}\tilde{G}^J = \tilde{\gamma}1$, $\tilde{\gamma} \neq 0$ in Ω. In particular for A of $\mathfrak{S}_J$ we have $A^G = \tilde{G}^{-1}A\tilde{G}$. Since the enveloping algebra of the matrices of $\mathfrak{S}_J$ is the algebra of matrices of $\mathfrak{A}$ we have $\tilde{G}^{-1}B\tilde{G}$ is a matrix of $\mathfrak{A}$ if B is and so $B \to \tilde{G}^{-1}B\tilde{G}$ is an automorphism of $\mathfrak{A}$. But the automorphisms of $\mathfrak{A}$ are inner and hence there is a matrix G of $\mathfrak{A}$ such that $\tilde{G}^{-1}B\tilde{G} = G^{-1}BG$. It follows that $\tilde{G} = \tilde{\rho}G$ and since $\tilde{G}$ is J-orthogonal, so is G.

THEOREM 4. *The group of automorphisms $\mathfrak{H}_J$ of $\mathfrak{S}_J \cong \mathfrak{G}_J/\mathfrak{D}$ where $\mathfrak{D}$ is the set of multiples $\delta 1$, $\delta \neq 0$ in Φ.*

Theorem 3 shows that $\mathfrak{G}_J$ is homomorphic to $\mathfrak{H}_J$. It is easily seen that the elements of $\mathfrak{G}_J$ which define the identity automorphism of $\mathfrak{S}_J$ are those of $\mathfrak{D}$. Hence $\mathfrak{H}_J \cong \mathfrak{G}_J/\mathfrak{D}$.

---

[20] Our proof breaks down for $\nu = 4$ since $\Lambda_1 = (\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4)/2$, $\Lambda_2 = (\lambda_1 + \lambda_2 - \lambda_3 - \lambda_4)/2$, $\Lambda_3 = (\lambda_1 - \lambda_2 + \lambda_3 - \lambda_4)/2$ and $\Lambda_4 = (\lambda_1 - \lambda_2 - \lambda_3 + \lambda_4)/2$ are weights satisfying our conditions and $\pm\Lambda_p \pm \Lambda_q$ give all the roots $\pm\lambda_i \pm \lambda_j$. The result does not hold in this case as has been shown by E. Cartan: *Le principe de dualité*, Bull. Sci. Math. **49** (1925) p. 367.

[21] Weyl I, p. 281.

[22] This theorem gives an algebraic connection between $\mathfrak{S}_J$ and $\mathfrak{G}_J$. If Φ is the field of complex numbers an analytic connection between these systems is known by Lie's theory.

5. **Isomorphism of distinct $\mathfrak{S}_J$'s.** Let $\mathfrak{A}_1$ and $\mathfrak{A}_2$ be two normal simple algebras and J_1, J_2 i. a. a. in $\mathfrak{A}_1$ and $\mathfrak{A}_2$ respectively such that $\mathfrak{S}_{J_1} \cong \mathfrak{S}_{J_2}$. Then $\mathfrak{S}_{J_1\Omega} \cong \mathfrak{S}_{J_2\Omega}$ and since the algebras of type 2 that we are considering are not isomorphic to any of type 1, $\mathfrak{S}_{J_1}$ and $\mathfrak{S}_{J_2}$ have the same type and hence $\mathfrak{A}_1$ and $\mathfrak{A}_2$ have the same degree n. We have seen that $\mathfrak{A}_1$ and $\mathfrak{A}_2$ have representations in Ω_n such that the linear combinations in Ω of the matrices of $\mathfrak{S}_{J_1}$ or $\mathfrak{S}_{J_2}$ comprise the elements of one of the systems (1), (2a) or (2b) denoted as (*). The isomorphism $A_1 \to A_2$ defines an automorphism in (*) and hence by Theorem 3 $A_2 = \bar{G}^{-1}A_1\bar{G}$. Since the enveloping algebras of the matrices of $\mathfrak{S}_{J_1}$ and $\mathfrak{S}_{J_2}$ using linear combinations in Φ only are the matrices of $\mathfrak{A}_1$ and $\mathfrak{A}_2$, the correspondence $B_1 \to \bar{G}^{-1}B_1\bar{G} = B_2$ for any matrix B_1 of $\mathfrak{A}_1$ defines an isomorphism between $\mathfrak{A}_1$ and $\mathfrak{A}_2$. If we identify $\mathfrak{A}_1$ and $\mathfrak{A}_2$ and repeat the argument we see that there is an automorphism of the matrices of $\mathfrak{A}$, $B \to \bar{G}^{-1}B\bar{G}$ sending the J_1-skew matrices into the J_2-skew matrices. It follows from the corollary to Theorem 2 that J_1 and J_2 are cogredient.

THEOREM 5. *A necessary and sufficient condition that $\mathfrak{S}_{J_1} \cong \mathfrak{S}_{J_2}$ where J_1 and J_2 are i. a. a. in the normal simple algebras $\mathfrak{A}_1$ and $\mathfrak{A}_2$ respectively is that $\mathfrak{A}_1 \cong \mathfrak{A}_2$ and J_1 and J_2 be cogredient.*

6. **Connection with the theory of abstract Lie algebras.** We have seen that any normal simple algebra with an i. a. a. J leads to a simple Lie algebra which becomes one of the systems (*) when the field is sufficiently extended. In this section we shall prove the converse, that any Lie algebra $\mathfrak{L}$ that has the type (*) when the field is extended is isomorphic to an algebra $\mathfrak{S}_J$.

Let $\mathfrak{L}$ be a Lie algebra over Φ and Ω the algebraic closure of Φ. We suppose that $\mathfrak{L}_\Omega$ has one of the types (*). Thus if $a_1, a_2, \cdots, a_m$ is a basis for $\mathfrak{L}$ and hence also for $\mathfrak{L}_\Omega$ there exist elements h, e_α linear combinations of the a_i in Ω satisfying the relations (2). But these linear combinations involve only a finite number of elements of Ω and so there is an algebraic extension $\bar{\Phi}$ of finite degree over Φ such that $\bar{\mathfrak{L}} = \mathfrak{L}_{\bar{\Phi}}$ has the type (*). We may suppose also that $\bar{\Phi}$ is a Galois field with the Galois group $\mathfrak{G} = (i, s, t \cdots)$. Thus $\mathfrak{L}$ has a representation by matrices in $\bar{\Phi}_n$ such that the linear combination of these matrices with coefficients in $\bar{\Phi}$ constitute the system (*). The matrices A_i corresponding to a_i form a basis for (*) over Φ and if $[a_i, a_j] = \sum \gamma_{ijk}a_k$ $(\gamma_{ijk} \in \Phi)$ then $[A_i, A_j] = \sum \gamma_{ijk}A_k$. It is clear from the definition of the system (*) that if $\bar{A} = (\bar{\alpha}_{ij})$ belongs to the system then so does $\bar{A}^s = (\alpha_{ij}^s)$ for any $s \in \mathfrak{G}$. In particular A_i^s belongs to the system, $A_1^s, \cdots, A_m^s$ are linearly independent relative to $\bar{\Phi}$ and $[A_i^s, A_j^s] = \sum \gamma_{ijk}A_k^s$ since $\gamma_{ijk}^s = \gamma_{ijk} \in \Phi$. It follows that for each $s \in \mathfrak{G}$ the correspondence $\bar{A} = \sum \bar{\alpha}_i A_i \to \bar{A}^S = \sum \bar{\alpha}_i A_i^s$ is an automorphism of the Lie algebra (*). By Theorem 3 there exists a matrix $\bar{G}_s$ in $\bar{\Phi}_n$ such that $\bar{A}^S = G_s^{-1}\bar{A}G_s$. For the matrices $A = \sum \alpha_i A_i$ $(\alpha_i \in \Phi)$ representing the elements of $\mathfrak{L}$ we have $A^s = A^S$ and hence

$$A^s = G_s^{-1}AG_s \qquad s \in \mathfrak{G}. \tag{6}$$

The totality of matrices satisfying the equations (6) is an algebra $\mathfrak{A}$ over Φ containing the matrices of $\mathfrak{L}$. Since the enveloping algebra of the matrices of $\tilde{\mathfrak{L}}$ is $\tilde{\Phi}_n$ there are n^2 linearly independent matrices $A_1, \cdots, A_{n^2}$ of the form $A_{i_1}A_{i_2}\cdots A_{i_k}$ $(i_\beta = 1, \cdots, m)$ and these belong to $\mathfrak{A}$. Now suppose $\tilde{B} = \sum_1^n \tilde{\beta}_i A_i$ is in $\mathfrak{A}$. Then

$$\sum \beta_i^s A_i^s = \tilde{B}^s = \tilde{G}_s^{-1}\tilde{B}\tilde{G}_s = \sum \tilde{\beta}_i \tilde{G}_s^{-1} A_i \tilde{G}_s = \sum \tilde{\beta}_i A_i^s$$

and hence $\tilde{\beta}_i^s = \tilde{\beta}_i = \beta_i \epsilon \Phi$. Thus $\mathfrak{A}$ consists of the totality of linear sums of $A_1, \cdots, A_{n^2}$ with coefficients in Φ, or $\mathfrak{A}$ is the enveloping algebra of the matrices of $\mathfrak{L}$ using coefficients in Φ only. Since $\tilde{\mathfrak{A}} = \mathfrak{A}_{\tilde{\Phi}} = \tilde{\Phi}_n$, $\mathfrak{A}$ is normal simple. Since $(A_{i_1}A_{i_2}\cdots A_{i_k})^J = (-1)^k A_{i_k}\cdots A_{i_2}A_{i_1} \epsilon \mathfrak{A}$ $(i_\beta = 1, 2, \cdots, m)$, $\mathfrak{A}$ is carried into itself by the i. a. a. J defined in $\tilde{\mathfrak{A}}$ and hence J is an i. a. a. in $\mathfrak{A}$. The matrices of $\mathfrak{L}$ are precisely the J-skew elements of $\mathfrak{A}$. We have therefore proved:

THEOREM 6. *If $\mathfrak{L}$ is a Lie algebra such that $\mathfrak{L}_\Omega$ is isomorphic to one of the algebras* (1), (2a) *or* (2b) *then $\mathfrak{L}$ may be realized as the set of J-skew elements of a normal simple algebra $\mathfrak{A}$.*

7. **Cogredience in a normal simple algebra.** It has been shown by Albert[23] that if $\mathfrak{A}$ is an involutorial normal simple algebra, i.e. has an i. a. a. defined in it then $\mathfrak{A} = \mathfrak{F}_m$ where $\mathfrak{F}$ is an involutorial normal division algebra. Thus $\mathfrak{F}$ has exponent 2 and its order is a power of 2 and if Φ is suitably restricted (e.g. an algebraic field of the rationals, a p-adic field, or the real field) then $\mathfrak{F}$ is a quaternion algebra.[24] We consider, however, the general case of Φ arbitrary and shall investigate the condition that two i. a. a.'s in $\mathfrak{F}_n$ be cogredient. The discussion is similar to that given in the proof of Theorem 1.

We begin with a fixed i. a. a. $A = (a_{ij}) \to (\bar{a}_{ji}) = \bar{A}'$ where $a_{ij} \epsilon \mathfrak{F}$ and $a \to \bar{a}$ is an i. a. a. in $\mathfrak{F}$. As before an arbitrary i. a. a. J in $\mathfrak{F}_n$ has the form $A^J = S^{-1}\bar{A}'S$ where $\bar{S}' = \pm S$, i.e. S is a hermitian or quasi-hermitian matrix. If $A^K = T^{-1}\bar{A}'T$ we have as before that K is cogredient to J if and only if $\tau T = \bar{V}'SV$, $\tau \epsilon \Phi$. Thus the question of cogredience of i. a. a. is one of ordinary cogredience of hermitian or quasi-hermitian matrices with elements in a normal division algebra. We hope to discuss this problem in a later paper.

UNIVERSITY OF CHICAGO.

---

[23] Loc. cit. in 10, pp. 894–901.

[24] These results are due to Brauer, Albert, Hasse and E. Noether. For proofs and references see Deuring's *Algebren*.

ANNALS OF MATHEMATICS
Vol. 38, No. 2, April, 1937

EINE LIMESBEZIEHUNG BEI GEWISSER VARIATION DER GRADZAHL EINES POLYNOMS

VON ERIK L. PETTERSON

(Received September 8, 1936)

Folgender Satz wird bewiesen:

SATZ. *Gegeben sei ein Polynom $f(x)$ der Form*

$$f(x) = g(x)^n\, M(x) + N(x)$$

mit festen normierten ganzzahligen Polynomen $g(x)$ und $M(x)$ und festem ganzzahligem Polynom $N(x)$. Ferner soll $N(x)$ keinen Faktor mit $g(x)\cdot M(x)$ gemeinsam haben. Von den Gradzahlen der irreduziblen Faktoren von $f(x)$, die nicht Teiler von einem Polynom der Form

$$g(x)^t - 1, \qquad (t \geqq 1),$$

sind, sei r_n die kleinste. Dann gilt

$$\lim_{n=\infty} r_n = \infty.$$

BEWEIS. Es sei ξ_n eine Wurzel von $f(x)$. Wird ξ_n immer so gewählt, dass

$$| g(\xi_n) | > 1 + \epsilon$$

ist, mit festem $\epsilon > 0$, wenn n eine unendliche Reihe von wachsenden positiven ganzen Zahlen durchläuft, so muss für diese n gelten

$$M(\xi_n) \to 0 \text{ für } n \to \infty. \tag{1}$$

Aus

$$| g(\xi_n) | < 1 - \epsilon$$

folgt in ähnlicher Weise

$$N(\xi_n) \to 0 \text{ für } n \to \infty. \tag{2}$$

Nach den Beziehungen (1) und (2) und der übrigen Möglichkeit

$$| g(\xi_n) | \to 1 \text{ für } n \to \infty$$

folgt, dass die Wurzeln von $f(x)$ für alle n beschränkt sind.

Ist die in dem Satz definierte Gradzahl r_n beschränkt, so muss wenigstens ein irreduzibler Faktor $A(x)$ vom konstanten Grade r für unendlich viele n vorkommen. Für hinreichend grosse n ist $f(x)$ und folglich auch $A(x)$ normiert, und man kann demnach setzen:

$$A(x) = x^r + b_1(n)x^{r-1} + b_2(n)x^{r-2} + \cdots + b_r(n).$$

Da die Wurzeln von $A(x)$ beschränkt sind, so sind die Koeffizienten $b_\nu(n)$, ($\nu = 1, 2, \cdots, r$), auch beschränkt. Diese $b_\nu(n)$ sind ferner ganze Zahlne, und es folgt dann, dass die Anzahl der Faktoren $A(x)$ des konstanten Grades r endlich ist. Es gibt daher wenigstens einen festen Faktor

$$A(x) = x^r + b_1x^{r-1} + \cdots + b_r$$

der für unendlich viele und also sicher für zwei verschiedene n vorkommt. Man erhält also

$$g(x)^{n_1}M(x) + N(x) = A(x)B_1(x)$$

$$g(x)^{n_2}M(x) + N(x) = A(x)B_2(x),\ n_2 > n_1,$$

mit ganzzahligen Polynomen $B_1(x)$ und $B_2(x)$. Nach Substraktion folgt

(3) $$g(x)^{n_1}M(x)(g(x)^{n_2-n_1} - 1) = A(x)(B_2(x) - B_1(x)).$$

Nach der Voraussetzung des Satzes enthalten $N(x)$ und $g(x)M(x)$ keinen gemeinsamen Faktor, und es gilt dann auch

$$A(x) \nmid g(x)M(x).$$

Aus (3) folgt demnach

$$A(x) \mid (g(x)^{n_2-n_1} - 1),$$

im Wilderspruch zur Voraussetzung, dass $A(x)$ nicht Teiler eines Polynoms der Form

$$g(x)^t - 1$$

sein soll.

Aus dem eben bewiesenen Satz folgt, dass jeder positiven ganzen Zahl v eine positive ganze Zahl U_v entspricht derart, dass $f(x)$, wenn

$$n \geqq U_v$$

ist, nur dann irreduzible Faktoren vom Grade $\leqq v$ enthalten kann, wenn diese Faktoren gleichzeitig Faktoren von Polynomen der Form $g(x)^t - 1$ sind. (Eine einfache Folgerung ist, dass diese Faktoren dann sogar für unendlich viele n vorkommt, nämlich für alle n aus einer arithmetischen Reihe $n = n_0 + it$, $i = 0, 1, 2, \cdots$).

Ist das Polynom

$$f(x) = g(x)^nM(x) + N(x)$$

entweder irreduzibel oder durch wenigstens einen Faktor teilbar, dessen Grad beschränkt bleibt, wenn n alle ganzen Zahlen > 0 durchläuft, und kann $f(x)$ für wenigstens hinreichend grosse n keinen Faktor mit Polynomen der Form

$g(x)^t - 1$ gemeinsam haben, so muss $f(x)$ irreduzibel sein für alle hinreichend grossen n. Wird daher beispielsweise die Theorie der Newtonschen Polygone nach Untersuchungen von Dumas und Ore[1] auf das Polynom $f(x)$ angewandt, so kann Irreduzibilität für alle hinreichend grossen n festgestellt werden. Werden

$$g(x) = x, \quad M(x) = 1 \text{ und } N(x) = ph(x)$$

gesetzt mit ganzzahligem Polynom $h(x)$, so erhält man folgendes Beispiel:

SATZ. *Im Polynom*

$$f(x) = x^n + ph(x), \qquad h(0) \neq 0,$$

seien p eine Primzahl und wenigstens ein Koeffizient in $h(x)$ nicht durch p teilbar. Dann gibt es eine feste positive Zahl U derart, dass $f(x)$ irreduzibel ist für alle $n > U$.

Dass $x^n + ph(x)$ keinen Faktor mit einem Polynom der Form $x^t - 1$ gemeinsam haben kann, wird folgendermassen gezeigt. Es seien ξ_ν für $\nu = 1, 2, \cdots, r$ die Wurzeln eines Faktors von $x^n + ph(x)$. Dann folgt

$$\left(\prod_{\nu=1}^{r} \xi_\nu\right)^n = \pm p^r \prod_{\nu=1}^{r} h(\xi_\nu)$$

und also

$$p \,\Big|\, \prod_{\nu=1}^{r} \xi_\nu$$

Sind diese ξ_ν auch Wurzeln von $x^t - 1$, so erhält man als Widerspruch

$$\left(\prod_{\nu=1}^{r} \xi_\nu\right)^t = 1$$

und

$$\prod_{\nu=1}^{r} \xi_\nu = \pm 1.$$

STOCKHOLM.

[1] Ore: "Zur Theorie der Irreduzibilitätskriterien," Math. Zeitschr. 18, S. 278–288, (1923).

ANNALS OF MATHEMATICS
Vol. 38, No. 2, April, 1937

ON SOME ANALYTIC SETS DEFINED BY TRANSFINITE INDUCTION

BY C. KURATOWSKI AND J. v. NEUMANN

(Received September 6, 1936)

In recent papers C. Kuratowski[1] has shown that—under very general conditions—transfinite induction applied in the domain of projective sets (in the sense of N. Lusin) does not lead outside of this domain. The purpose of this paper is to prove a theorem which gives under some supplementary assumptions a more precise evaluation of the projective class of a considered set. Thus the set of Lebesgue[1a] which was shown by the method of Kuratowski to be of class 3 will be seen to be a difference of two analytic sets (hence of class 2).

1. Let $r_1, r_2, \cdots$ be an enumeration of the set of all rational numbers, which will be held fixed in what follows. Let C be the Cantor perfect set of all

$$x = \frac{c_1}{3} + \frac{c_2}{9} + \cdots \qquad (c_1 = 0 \text{ or } 2).$$

For any $x \in C$ form the set M_x of all r_i with $c_i = 2$. Put $\bar{x}$ = ordinal type of M_x in its natural ordering (and not that one of the enumeration $r_1, r_2, \cdots$). Thus $\bar{x}$ runs over all ordinal enumerable types, if x runs over C. Denote by C_0 the set of all numbers x such that M_x is a well-ordered set (i.e. $\bar{x} < \Omega$). Observe that $x \rightleftarrows M_x$ is a one-to-one correspondence of all elements of C and all sets of rational numbers.

2. The following notations will be used in the definition of the set of Lebesgue. For any $x \in C$ put

$$x_n = \frac{c_{1\cdot 2^n}}{3} + \frac{c_{3\cdot 2^n}}{9} + \frac{c_{5\cdot 2^n}}{27} + \cdots \qquad (n = 0, 1, 2, \cdots)$$

The symbol $x^{(n)}$ is defined as follows: if M_x contains an $r_k \geqq r_n$ then $M_{x^{(n)}}$ is composed of all r_i such that $r_i < r_n$ and $r_i \in M_x$; if not, put $x^{(n)} = 0$. Thus, if $x \in C_0$ and $x \neq 0$, $\overline{x^{(n)}}$ runs over all ordinal numbers less than $\bar{x}$.

Let $A(y)$, $y \in C$, be a "universal" function relatively to the family of closed and of open subsets of the interval.[2] We may assume that the plane set $E_{yz}[z \in A(y)]$ is a Borel set.

Now define for any xy with $x \in C_0$ and $y \in C$ the set $L(x, y)$ as follows:

---

[1] C. R. Paris t. 202 (1936), p. 1239 and Fund. Math. 27 (1936).
[1a] Journal de Mathématiques, 1905, Chap. VIII.
[2] This means that X being any set belonging to the family considered, there exists a y such that $X = A(y)$.

$g(x)^t - 1$ gemeinsam haben, so muss $f(x)$ irreduzibel sein für alle hinreichend grossen n. Wird daher beispielsweise die Theorie der Newtonschen Polygone nach Untersuchungen von Dumas und Ore[1] auf das Polynom $f(x)$ angewandt, so kann Irreduzibilität für alle hinreichend grossen n festgestellt werden. Werden

$$g(x) = x, \quad M(x) = 1 \text{ und } N(x) = ph(x)$$

gesetzt mit ganzzahligem Polynom $h(x)$, so erhält man folgendes Beispiel:

SATZ. *Im Polynom*

$$f(x) = x^n + ph(x), \qquad h(0) \neq 0,$$

seien p eine Primzahl und wenigstens ein Koeffizient in $h(x)$ nicht durch p teilbar. Dann gibt es eine feste positive Zahl U derart, dass $f(x)$ irreduzibel ist für alle $n > U$.

Dass $x^n + ph(x)$ keinen Faktor mit einem Polynom der Form $x^t - 1$ gemeinsam haben kann, wird folgendermassen gezeigt. Es seien ξ_ν für $\nu = 1, 2, \cdots, r$ die Wurzeln eines Faktors von $x^n + ph(x)$. Dann folgt

$$\left(\prod_{\nu=1}^{r} \xi_\nu\right)^n = \pm p^r \prod_{\nu=1}^{r} h(\xi_\nu)$$

und also

$$p \,\Big|\, \prod_{\nu=1}^{r} \xi_\nu$$

Sind diese ξ_ν auch Wurzeln von $x^t - 1$, so erhält man als Widerspruch

$$\left(\prod_{\nu=1}^{r} \xi_\nu\right)^t = 1$$

und

$$\prod_{\nu=1}^{r} \xi_\nu = \pm 1.$$

STOCKHOLM.

---

[1] Ore: "Zur Theorie der Irreduzibilitätskriterien," Math. Zeitschr. 18, S. 278–288, (1923).

ANNALS OF MATHEMATICS
Vol. 38, No. 2, April, 1937

ON SOME ANALYTIC SETS DEFINED BY TRANSFINITE INDUCTION

BY C. KURATOWSKI AND J. V. NEUMANN

(Received September 6, 1936)

In recent papers C. Kuratowski[1] has shown that—under very general conditions—transfinite induction applied in the domain of projective sets (in the sense of N. Lusin) does not lead outside of this domain. The purpose of this paper is to prove a theorem which gives under some supplementary assumptions a more precise evaluation of the projective class of a considered set. Thus the set of Lebesgue[1a] which was shown by the method of Kuratowski to be of class 3 will be seen to be a difference of two analytic sets (hence of class 2).

1. Let $r_1, r_2, \cdots$ be an enumeration of the set of all rational numbers, which will be held fixed in what follows. Let C be the Cantor perfect set of all

$$x = \frac{c_1}{3} + \frac{c_2}{9} + \cdots \qquad (c_i = 0 \text{ or } 2).$$

For any $x \in C$ form the set M_x of all r_i with $c_i = 2$. Put $\bar{x}$ = ordinal type of M_x in its natural ordering (and not that one of the enumeration $r_1, r_2, \cdots$). Thus $\bar{x}$ runs over all ordinal enumerable types, if x runs over C. Denote by C_0 the set of all numbers x such that M_x is a well-ordered set (i.e. $\bar{x} < \Omega$). Observe that $x \rightleftarrows M_x$ is a one-to-one correspondence of all elements of C and all sets of rational numbers.

2. The following notations will be used in the definition of the set of Lebesgue. For any $x \in C$ put

$$x_n = \frac{c_{1\cdot 2^n}}{3} + \frac{c_{3\cdot 2^n}}{9} + \frac{c_{5\cdot 2^n}}{27} + \cdots \qquad (n = 0, 1, 2, \cdots)$$

The symbol $x^{(n)}$ is defined as follows: if M_x contains an $r_k \geqq r_n$ then $M_{x^{(n)}}$ is composed of all r_i such that $r_i < r_n$ and $r_i \in M_x$; if not, put $x^{(n)} = 0$. Thus, if $x \in C_0$ and $x \neq 0$, $\overline{x^{(n)}}$ runs over all ordinal numbers less than $\bar{x}$.

Let $A(y)$, $y \in C$, be a "universal" function relatively to the family of closed and of open subsets of the interval.[2] We may assume that the plane set $E_{yz}[z \in A(y)]$ is a Borel set.

Now define for any xy with $x \in C_0$ and $y \in C$ the set $L(x, y)$ as follows:

---

[1] C. R. Paris t. 202 (1936), p. 1239 and Fund. Math. 27 (1936).

[1a] Journal de Mathématiques, 1905, Chap. VIII.

[2] This means that X being any set belonging to the family considered, there exists a y such that $X = A(y)$.

$$\begin{cases} L(0, y) = A(y) \\ L(x, y) = \operatorname*{Lim\,sup}_{n\to\infty} L(x^{(n)}, y_n) \qquad \text{for } x \neq 0^{3}. \end{cases}$$

The spatial set $T = E_{xyz}[z \in L(x, y)]$ is called the Lebesgue set.

3. The preceding construction of the set T may be generalized as follows. Replace the function $x^{(n)}$ by any function $f_n(x)$ such that $f_n(0) = 0$ and for $0 \neq x \in C_0$, $\overline{f_n(x)} < \bar{x}$. Let y_n be replaced by an arbitrary function $g_n(y)$ of y and let $A(y)$ denote an arbitrary set-function. The operation Lim sup will be replaced by a general Hausdorff operation.

Let us recall the definition of a Hausdorff operation. Let B be a set of sequences composed of positive integers. Denote by $s = (s^1, s^2, \cdots)$ a variable sequence. Given a sequence of sets $X_1, X_2, \cdots$, form the set

$$\sum_s \prod_n X_{s^n}, \text{ where } s \text{ ranges over } B.$$

This set is said to be obtained with the aid of the Hausdorff operation with base B.

In what follows the sequence s will be identified with the irrational number $s = \frac{1|}{|s^1} + \frac{1|}{|s^2} + \cdots$. Thus B will become a subset of the set of all irrational numbers. (E.g. in the case of the operation Lim sup the set B is composed of all numbers s containing infinitely many different s^n).

Given the functions $f_n(x)$, $g_n(y)$, $A(y)$ and the base B, it is readily proved with the aid of transfinite induction that there exists one and only one function $L(x, y)$ defined for all $x \in C_0$ and such that

$$\text{(1)} \qquad \begin{cases} L(0, y) = A(y) \\ L(x, y) = \sum_s \prod_n L[f_{s^n}(x), g_{s^n}(y)] \text{ where } s \in B \text{ and } x \neq 0. \end{cases}$$

THEOREM. *If $f_n(x)$ and $g_n(y)$ are Baire functions and B and $E_{yz}[z \in A(y)]$ are analytic sets, the spatial set $T = E_{xyz}[z \in L(x, y)]$ is an analytic set relatively to the part of the space composed of points xyz such that $x \in C_0$; i.e. T is the intersection of an analytic set U and of the product $C_0 \times Y \times Z$.*

4. The following notation will be used in the proof of the theorem.

Given a correspondence which assigns to each finite system of positive integers $k_1, k_2, \cdots, k_n$ $(n \geqq 0)$ a non-negative integer $p_{k_1 \ldots k_n}$, denote by P the infinite system $(p, p_1, p_2, \cdots, p_{k_1 \cdots k_n}, \cdots)$. Obviously the set $\mathfrak{P}$ of all P's may be considered as topologically equivalent to the space of all infinite sequences of non-negative integers, hence to the space of all irrational numbers. This allows

[3] Lim sup $X_n = \prod_{n=0}^{\infty} \sum_{k=0}^{\infty} X_{n+k}$.

us to apply theorems concerning the use of the logical quantifiers $\sum_P$ and $\prod_P$[4] with P as variable: so if $\varphi(P, x)$ is a given relation between the variables P and x, the set $E_x \sum_P \varphi(P, x)$ is the projection eliminating the variable P of the "plane" set $E_{xP}\, \varphi(P, x)$; hence if the last set is analytic, so is the first.

Denote by $S_{k_1 \cdots k_n}(P)$ the sequence $(p_{k_1 \cdots k_n 1}, p_{k_1 \cdots k_n 2}, \cdots)$. Thus $S(P) = (p_1, p_2, \cdots)$.

It will be observed that, if $k_1, \cdots, k_n$ is given, then $p_{k_1 \cdots k_n}$ and $S_{k_1 \cdots k_n}(P)$ are continuous functions of P.

Finally let us write $x^n = f_n(x)$, $x^{k_1 \cdots k_n} = f_{k_n}(\cdots f_{k_1}(x) \cdots)$ and $y_n = g_n(y)$ and $y_{k_1 \cdots k_n} = g_{k_n}(\cdots g_{k_1}(y) \cdots)$. If the functions $f_n(x)$ and $g_n(y)$ are Baire functions, the same holds true of the functions $x^{k_1 \cdots k_n}$ and $y_{k_1 \cdots k_n}$.

Proof of the theorem. Denote by H any finite system of positive integers $k_1 \cdots k_n$ $(n \geqq 0)$ and put $| P, H | = (p_{k_1}, p_{k_1 k_2}, \cdots p_{k_1 k_2 \cdots k_n})$. If $n = 0$ we have $| P, H | = 0$ (the vacuous set), $p_H = p$ and $x^{|P,H|} = x$, $y_{|P,H|} = y$.

Consider the set U of all points xyz $(x \in C)$ satisfying the following condition: there exists a $P = (p, p_1, p_2, \cdots p_H, \cdots)$ such that:

1^0: $p \neq 0$

2^0: if $p_H \neq 0$ and $x^{|P,H|} = 0$, then $z \in A(y_{|P,H|})$

3^0: if $p_H \neq 0$ and $x^{|P,H|} \neq 0$, then $S_H(P) \in B$.

In logical symbols:

$$(2) \quad \begin{cases} (xyz \in U) \equiv \sum_P \Big\{ (p \neq 0) \prod_H [(p_H \neq 0)(x^{|P,H|} = 0) \rightarrow z \in A(y_{|P,H|}) \\ \qquad\qquad\qquad\qquad (p_H \neq 0)(x^{|P,H|} \neq 0) \rightarrow S_H(P) \in B] \Big\}. \end{cases}$$

The hypotheses of the theorem lead easily to the conclusion that the set of points $(xyzP)$ satisfying the condition in brackets $\{\}$ is an analytic set (in the space $C \times Y \times Z \times \mathfrak{P}$). Therefore the projection of that set, i.e. the set U, is an analytic set too. It remains to show that, assuming $x \in C_0$, U may be replaced in the preceding formula by T.

1) Assume that $(xyz) \in T$, i.e. that $z \in L(x, y)$. We shall show by transfinite induction that there exists a P verifying the conditions 1^0–3^0 and such that p = any preassigned value.

Consider first the case $x = 0$. Then by (1) $z \in A(y)$. Define P as follows: p = any number $\neq 0$ and for $H \neq 0$ put $p_H = 0$. The conditions 1^0–3^0 are obviously satisfied.

Now let $\alpha > 0$ and suppose our assumption holds true for any x such that $\bar{x} < \alpha$.

Let $\bar{x} = \alpha$. Since $(xyz) \in T$, it follows by (1) that there exists a sequence

---

[4] $\sum_x$ means "there exists an x such that $\cdots$" and $\prod_x$ means "for any $x \cdots$" Concering the properties of the quantifiers, see e.g. C. Kuratowski *Topologie* I, Warsaw 1933, §1-2.

of positive integers $s = (s^1, s^2, \cdots) \in B$ such that, for any n, $z \in L[f_{s^n}(x), g_{s^n}(y)]$. That means that $(x^{s^n}, y_{s^n}, z) \in T$ and since $\overline{x^{s^n}} < \bar{x}$, there exists a

$$P^n = (p^n, p_1^n, \cdots p_H^n, \cdots)$$

such that:

1_n^0: $p^n = s^n$

2_n^0: if $p_H^n \neq 0$ and $x^{s^n | P^n, H|} = 0$, then $z \in A(y_{s^n|P^n,H|})$

3^0: if $p_H^n \neq 0$ and $x^{s^n|P^n,H|} \neq 0$, then $S_H(P^n) \in B$.

Define P as follows: $p =$ any number $\neq 0$ and for $n \neq 0$ put $p_{k_1 \cdots k_n} = p_{k_2 \cdots k_n}^{k_1}$. Then $s^n \mid P^n, H \mid = p_n \mid P^n, H \mid = \mid P, nH \mid$, $S_H(P^n) = S_{nH}(P)$ and it is easily seen that the conditions $1^0 - 3^0$ are satisfied.

2) Now assume that P satisfies the conditions $1^0 - 3^0$. We have to show that $(xyz) \in T$, i.e. that $z \in L(x, y)$.

This is obvious in the case $x = 0$. For, the condition 2^0 gives, for $H = 0$, $z \in A(y)$, hence, by (1): $z \in L(0, y)$.

Let $\alpha > 0$ and suppose our assumption is true for x's such that $\bar{x} < \alpha$.

Consider an x such that $\bar{x} = \alpha$ and a P satisfying the conditions $1^0 - 3^0$.

Put $s = (p_1, p_2, \cdots)$ i.e. $s = S(P)$. Then the conditions 1^0 and 3^0 give, for $H = 0$, $S(P) \in B$. Thus it remains to show that $z \in L[f_{s^n}(x), g_{s^n}(y)]$. But this follows easily from the fact that if P^n be defined by putting $p_{k_1 \cdots k_m}^n = p_{nk_1 \cdots k_m}$, then the conditions $1_n^0 - 3_n^0$ are fulfilled.

Thus the theorem is proved. It follows at once that the Lebesgue set is the difference of two analytic sets (the set C_0 being the complement of an analytic set).[5]

5. **Some generalizations.** It is readily seen that the proof of the theorem is not affected if we assume that Y and Z denote any complete separable spaces, that f_n and g_n are Baire functions of both variables x and y, that B also depends upon x and y and that the set $E_{xys}[s \in B_{x,y}]$ is analytic.

Thus we shall have to replace (1) by

$$(3) \quad \begin{cases} L(0, y) = A(y) \\ L(x, y) = \sum_s \prod_n L[f_{s^n}(x, y), g_{s^n}(x, y)] \text{ where } s \in B_{x,y} \text{ and } x \neq 0. \end{cases}$$

On the right hand side of (2) we replace x and y by (x, y) and B by

$$B_{(x,y)|P,H|,(x,y)|P,H|}.^{6}$$

This generalized theorem may be applied to some other definitions based on transfinite induction.

---

[5] See e.g. *Topologie* I p. 257.

[6] Thus we have $(x, y)^n = f_n(x, y)$, $(x, y)_n = g_n(x, y)$, $(x, y)^{n,m} = f_m[(x, y)^n, (x, y)_n]$ etc. For $n = 0$ we have $(x, y)^{k_1 \cdots k_n} = x$ and $(x, y)_{k_1 \cdots k_n} = y$

1) Consider first the case where $Y = Z = C$, where $A(y)$ denotes a universal function for the open subsets of C such that the set $\underset{yz}{E}\,[z \in A(y)]$ is open and where x_n and $x^{(n)}$ are defined as in §2. Put, for $x \in C_0$,

$$\begin{cases} L(0, y) = A(y) & \\ L(x, y) = \prod_n L(x^{(n)}, y_n) & \text{for } \bar{x} \text{ odd} \\ = \sum_n L(x^{(n)}, y_n) & \text{for } \bar{x} \text{ even} \neq 0. \end{cases}$$

The fundamental property of the function $L(x, y)$ is that, for any $x \in C_0$, $L(x, y)$ is a universal function for Borel sets of class $G_{\bar{x}}$ and the set $\underset{yz}{E}\,[z \in L(x, y)]$ is of that class.[7]

Now here the base B depends on x: for $\bar{x}$ odd it is composed of one sequence, namely of the sequence of all positive integers, and for $\bar{x}$ even B is the set of all sequences of positive integers. Since the set of all x's (belonging to C) such that $\bar{x}$ is odd is a Borel set,[8] it follows easily that the set $\underset{xs}{E}\,[s \in B_x]$ is a Borel set. Hence the corresponding set U is an analytic set and the set $T = \underset{xyz}{E}\,[z \in L(x, y)]$ is a difference of two analytic sets.

2) Using the notations of the preceding example, define x^*, for $\bar{x}$ odd, so that the set M_{x^*} is composed of all elements of M_x except the last one. Thus $\bar{x} = x^* + 1$. Put[9]

$$\begin{cases} L(0, y) = A(y) \\ L(x, y) = \prod_n L(x^*, y_n) \text{ for } \bar{x} \text{ odd, and for } \bar{x} \text{ even, } \neq 0 \text{, put} \\ L(x, y) = \sum_n L(y_{2n+1}, y_{2n}) \text{ where } n \text{ ranges over indices such that } \bar{y}_{2n+1} < \bar{x}. \end{cases}$$

In other terms, define the functions f_n and g_n as follows:

(i) if $\bar{x}$ is odd, $f_n(x, y) = x^*$ and $g_n(x, y) = y_n$,

(ii) if $\bar{x}$ is even, $f_n(x, y) = y_{2n+1}$ and $g_n(x, y) = y_{2n}$ and let $B_{x,y}$ be defined for $\bar{x}$ odd as in the preceding example and for $\bar{x}$ even suppose that it is composed of all sequences s such that for any n we have $\bar{y}_{2s^n+1} + 1 \leqq \bar{x}$. The set

$$\underset{xys}{E}\,[s \in B_{x,y}] = \prod_n \underset{xys}{E}\,[\bar{y}_{2s^n+1} + 1 \leqq \bar{x}]$$

being analytic,[10] the corresponding set U is analytic and T is a difference of two analytic sets.

INSTITUTE FOR ADVANCED STUDY.

---

[7] See e.g. *Topologie* I, p. 173.

[8] $\bar{x}$ may be called *odd* (for any $x \in C$) if it is of the form $\lambda + 2n + 1$ where λ is an ordinal type without last element. See Fund. Math. 27 op. cit.

[9] The essential difference between the examples 1) and 2) is that in the second case the function $L(x, y)$, for a given y, depends upon $\bar{x}$ alone (and not upon x). See C. Kuratowski, C. R. Paris t. 176 (1923) p. 229 and *Topologie* I, p. 175.

[10] By a theorem of Lusin the set of xy such that $\bar{x} \leqq \bar{y}$ is analytic. See e.g. *Topologie* I, p. 258.

ANNALS OF MATHEMATICS
Vol. 38, No. 2, April, 1937

REPRESENTABILITY OF LIE ALGEBRAS AND LIE GROUPS BY MATRICES

BY GARRETT BIRKHOFF

(Received September 14, 1936)

1. **Introduction.** Let A be any associative hypercomplex algebra over any commutative field F. If one defines in A *alternants* $[xy]$ by setting

$$[xy] = yx - xy$$

then with respect to the (generally non-associative) "products" $[xy]$, the elements of A form a new hypercomplex algebra $L(A)$ over F. Moreover it is easy to verify the well-known identities of Lie-Jacobi in $L(A)$, namely

$$[xy] + [yx] = 0 \qquad [[xy]z] + [[yz]x] + [[zx]y] = 0.$$

Hypercomplex algebras satisfying these identities are usually called "Lie algebras," and it is clear that every module (= linear subset) in A which contains the alternant of any two of its members, forms a Lie algebra.

The first main result proved below is that (Theorem 1) conversely every Lie algebra can be obtained in this way.

But the result is defective in that, although $L(A)$ is of finite order whenever A is, it is not proved that every Lie algebra L of finite order can be obtained from an associative algebra of finite order. In other words, it is not proved that every such L can be represented isomorphically by finite matrices. This problem is of course by Lie's Third Fundamental Theorem equivalent in the cases that F is the real resp. complex field to the classical question: is every Lie group nucleus G isomorphic locally[1] with a group nucleus of matrices?

The main result bearing on this problem which has been proved hitherto is due to Lie, and states that in case G has a discrete central, it is locally isomorphic with the group of linear transformations which its inner automorphisms induce on its infinitesimal elements (its so-called "adjoint" group). It is known[2] that this corresponds to the statement that any Lie algebra L of finite order which possesses no element a satisfying $[xa] = 0$ for all x can be obtained from an "adjoint" associative algebra of finite order.

The second new result proved below is the fact that (Theorem 4) in the

---

[1] It is known (cf. infra) that there exist Lie groups not isomorphic in the large with any group of finite matrices. The reader's attention is called to a paper of I. Ado (Russian) in the Bull. physico-math. society of Kazan 6 (1935), which contains the theorem that every Lie algebra over the real or complex field is isomorphic with a Lie algebra of matrices.

[2] N. Jacobson, "Rational methods in the theory of Lie algebras," Annals of Math. 36 (1935), 875–81.

extreme opposite case that L is *nilpotent,* it can be obtained from a nilpotent associative algebra of finite order. And using this it is proved that (Theorem 6) any "hypercentral" Lie group nucleus is locally isomorphic with a simply connected Lie group of matrices—whence the universal covering group of any "hypercentral" Lie group is topologically isomorphic with a Lie group of matrices.

2. **Canonical polynomials and "straightening".** Let L be any Lie algebra, and let us suppose that $e_1, e_2, e_3, \cdots$ constitute a basis of linearly independent elements of L, well-ordered in some fixed way.[3] Further, let the rule for commutation in L be expressed by the equations

$$[e_i e_j] = \sum_k c_k^{ij} e_k.$$

Since we are in the realm of pure algebra, all sums considered are finite.

Now suppose A is any associative algebra containing L and generated by the elements of L. Since A is generated by L, any element of A can be written as a polynomial—that is, linear combination of products $\xi = e_{i(1)} \cdots e_{i(q)}$—in the e_i.

Suppose now we call a "straightening" of any polynomial, a substitution for one of its terms $\xi = e_{i(1)} \cdots e_{i(q)}$ of the *equal* element

$$e_{i(1)} \cdots e_{i(j-1)} e_{i(j+1)} e_{i(j)} e_{i(j+2)} \cdots e_{i(q)} + \sum_k c_k^{i(j+1)\,i(j)} e_{i(1)} \cdots e_{i(j-1)} e_k e_{i(j+2)} \cdots e_{i(q)}$$

in case $i(j+1) < i(j)$.

It is clear that after at most $\frac{1}{2}q(q-1)$ straightenings, one can replace any such product ξ by an equal polynomial $\xi^* + \sum_k a_k \xi_k$, where ξ^* is a product of the same degree as ξ and has the "canonical" form

$$e_{h(1)}^{p1} \cdots e_{h(r)}^{pr} \qquad [h(1) < \cdots < h(r)]$$

and the ξ_k are products of lower degree. Hence, by induction

LEMMA 1: *After being straightened, any element of A can be shown to be equal to a linear combination of canonical power-products of the e_i.*

Observe that one can calculate the canonical polynomial equivalent to a given polynomial of A simply by knowing L; the nature of A is irrelevant. This suggests, given L, reducing symbolic polynomials in its basis-elements to canonical form without introducing A. But as this can be done in various ways, it is essential to know that

LEMMA 2: *The reduction of polynomials in the e_i through straightening to canonical form, is independent of the method of straightening.*

PROOF: Let ξ be any product $e_{i(1)} \cdots e_{i(q)}$, and let $\phi'(\xi)$ and $\phi''(\xi)$ be the results of two different straightenings of ξ. Then if we can find ways of straight-

[3] It follows from the axiom of choice that any module has such a basis: one need merely add to a partial basis successively new linearly independent elements as long as any exist.

ening $\phi'(\xi)$ and $\phi''(\xi)$ respectively into the same polynomial $\phi'''(\xi)$, we have shown by induction on the degree and the number of straightenings still possible of the product of highest degree, that *all* ways of straightening ξ lead to the same canonical polynomial

$$\gamma(\phi'(\xi)) = \gamma(\phi'''(\xi)) = \gamma(\phi''(\xi)).$$

This is what we shall do.

It is clear that there are essentially only two cases, which may be summarized in the two sets of formulas:

$$\text{I.}\quad \begin{cases} \xi = \xi_1 e_{i'} e_{j'} \xi_2 e_{i''} e_{j''} \xi_3 \\ \phi'(\xi) = \xi' + \eta' = \xi_1 e_{j'} e_{i'} \xi_2 e_{i''} e_{j''} \xi_3 + \sum_k c_k^{j'i'} \xi_1 e_k \xi_2 e_{i''} e_{j''} \xi_3 \\ \phi''(\xi) = \xi'' + \eta'' = \xi_1 e_{i'} e_{j'} \xi_2 e_{j''} e_{i''} \xi_3 + \sum_k c_k^{j''i''} \xi_1 e_{i'} e_{j'} \xi_2 e_k \xi_3. \end{cases}$$

$$\text{II.}\quad \begin{cases} \xi = \xi_1 e_h e_i e_j \xi_2 \\ \phi'(\xi) = \xi' + \eta' = \xi_1 e_i e_h e_j \xi_2 + \sum_k c_k^{ih} \xi_1 e_k e_j \xi_2 \\ \phi''(\xi) = \xi'' + \eta'' = \xi_1 e_h e_j e_i \xi_2 + \sum_k c_k^{ji} \xi_1 e_h e_k \xi_2. \end{cases}$$

In Case I, ξ' and ξ'' can be further straightened, thus

$$\xi' \to \xi''' + \eta_1'' = \xi_1 e_{j'} e_{i'} \xi_2 e_{j''} e_{i''} \xi_3 + \sum_k c_k^{j''i''} \xi_1 e_{j'} e_{i'} \xi_2 e_k \xi_3$$

$$\xi'' \to \xi''' + \eta_1' = \xi_1 e_{j'} e_{i'} \xi_2 e_{j''} e_{i''} \xi_3 + \sum_k c_k^{j'i'} \xi_1 e_k \xi_2 e_{j''} e_{i''} \xi_3.$$

And straightening the $e_{i''} e_{j''}$ resp. $e_{i'} e_{j'}$ terms in η' resp. η'', one finds that $\eta' + \eta_1''$ can be straightened into the same form as $\eta_1' + \eta''$—giving finally

$$\phi'''(\xi) = \mu' + \eta_1' + \eta_1'' + \sum_{k'} \sum_{k''} c_{k'}^{j'i'} c_{k''}^{j''i''} \xi_1 e_{k'} \xi_2 e_{k''} \xi_3.$$

Similarly in case II, evidently $h > i > j$, and so we can straighten further by setting

$$\begin{aligned} \xi' &\to \xi_1 e_i e_j e_h \xi_2 + \sum_k c_k^{jh} \xi_1 e_i e_k \xi_2 \\ &\to \xi_1 e_j e_i e_h \xi_2 + \sum_k c_k^{ji} \xi_1 e_k e_h \xi_2 + \sum_k c_k^{jh} \xi_1 e_i e_k \xi_2 \\ &= \xi''' + \eta_1'' + \eta_1'''. \end{aligned}$$

$$\begin{aligned} \xi'' &\to \xi_1 e_j e_h e_i \xi_2 + \sum_k c_k^{jh} \xi_1 e_k e_i \xi_2 \\ &\quad\ \xi_1 e_j e_i e_h \xi_2 + \sum_k c_k^{ih} \xi_1 e_j e_k \xi_2 + \sum_k c_k^{jh} \xi_1 e_k e_i \xi_2 \\ &= \xi''' + \eta_1' + \eta_2'''. \end{aligned}$$

Hence it is sufficient to show that $\eta' + \eta_1'' + \eta_2'''$ and $\eta_1' + \eta'' + \eta_2'''$ can be straightened into the same form. But if one transposes $e_k e_j$, $e_k e_i$, $e_k e_h$, $e_j e_k$, and $e_h e_k$ whenever they are in the wrong order, one finds that this assertion comes down precisely to Jacobi's identity, which completes the proof.

3. **Corollaries of Lemmas 1-2.** From Lemmas 1–2 one can deduce our first main result. For starting with a Lie algebra purely abstractly, taking the canonical polynomials in its basis-elements e_i as elements of a hypercomplex algebra A, multiplying them symbolically and straightening the products to canonical form through Lemma 1, it follows[4] from Lemma 2 that multiplication is associative. Hence since by the first identity of Lie-Jacobi and definition,

$$e_j e_i - e_i e_j = c_k^{ij} e_k .$$

THEOREM 1: *Any Lie algebra L can be imbedded in a linear associative algebra $A_u(L)$.*

COROLLARY: *Any Lie algebra is isomorphic with an algebra of infinite matrices, with respect to the operation of forming alternants* $[XY] = YX - XY$.

For if one adds to the basis of any linear associative algebra A a "principal unit" e satisfying $ex = xe = x$ for all x, then the correspondence between the elements x of A and the linear transformations X: $a \to ax = X(a)$ of A is *isomorphic* (it is in any case homomorphic). This well-known construction goes back to Poincaré.[5]

Now recall that if A is any linear associative algebra in which L is embedded isomorphically and which is generated by L, then the correspondence between polynomials in the basis-elements e_i of L and their values in A preserves all equalities true in $A_u(L)$. Hence

THEOREM 2: *The ways of imbedding a given Lie algebra L in a linear associative algebra A are all obtained by taking the invariant subalgebras S of $A_u(L)$ which place the different linear combinations of the e_i in different residue classes, and then forming the homomorphic images $A_u(L)/S$.*

Thus $A_u(L)$ is a kind of universal linear associative algebra containing L. Finally, we have as a corollary of Theorem 1,

THEOREM 3: *The free Lie algebra with n generators is isomorphic with the free algebra of alternants involving n symbols.*

In other words, the identities of Lie-Jacobi imply all all other identities true of alternants.

The author has been informed that much of §3 was also discovered, but not published, by E. Artin.

---

[4] Given monomials ξ_1, ξ_2 and ξ_3, one writes down $\xi_1 \xi_2 \xi_3$ purely symbolically; straightening first $\xi_1 \xi_2$ one gets $(\xi_1 \xi_2)\xi_3$, and straightening first $\xi_2 \xi_3$ one gets $\xi_1(\xi_2 \xi_3)$.

[5] L. E. Dickson, "Algebras and their arithmetics," Chicago, 1922, pp. 92–98 discusses it more elaborately.

4. The case of nilpotent Lie algebras. A Lie algebra L is called *nilpotent* if and only if all its brackets of sufficient length w vanish,[6] and a matrix $X = \| x_{ij} \|$ is called *triangular* if and only if the elements below the principal diagonal vanish (i.e., $i < j$ implies $x_{ij} = 0$), and *properly triangular* if and only if in addition the principal diagonal consists of zeros.

Suppose L is a nilpotent Lie algebra. Define $L_1 = L$, and L_k $[k = 2, 3, \cdots]$ inductively as the set of all brackets $[xy]$ with $x \in L_i$ and $y \in L_{k-i}$, and $0 < i < k$—that is, define L_k as the set of all complex brackets of length k. Further, let J_h $[h = 1, 2, 3, \cdots]$ denote the set of linear combinations of elements of $L_h, L_{h+1}, L_{h+2}, \cdots$—that is, of polynomials in brackets of length $\geqq h$. Then the J_h are invariant subalgebras of L, and by the hypothesis of nilpotence, for some finite integer w,

$$L = J_1 > J_2 > J_3 > \cdots > J_w = 0.$$

Evidently L possesses a basis of linearly independent elements $e_1, \cdots, e_r$ such that

$$\begin{array}{ll} e_1, \cdots, e_{n(1)} & \text{are a basis for } J_{w-1} \\ e_1, \cdots, e_{n(2)} & \text{are a basis for } J_{w-2} \\ \cdot \quad \cdot \quad \cdot & \cdot \quad \cdot \quad \cdot \quad \cdot \\ e_1, \cdots, e_{n(w-1)} & \text{are a basis for } J_1 = L. \end{array}$$

Moreover obviously $n(1) < n(2) < \cdots < n(w-1) = r$.

Now taking the particular basis $e_1, \cdots, e_r$, let us form $A_u(L)$ as in §3. And to each e_k let us ascribe a numerical "weight" $s = s(e_k)$, so chosen that e_k is in J_s but not in J_{s+1}. Further, to each product let us ascribe as "weight" the sum of the weights of its factors, and to each polynomial, the least of the weights of its monomial terms. Evidently (since if $[e_i e_j] = c_k^{ij} e_k$, then $s(k) < s(i) + s(j)$ implies $c_k^{ij} = 0$ by our choice of basis), straightening preserves lower bounds to the weights of the terms of any polynomial, and multiplying a polynomial by another has the same property.

Hence the polynomials of weight $> w$ constitute an *invariant subalgebra* S of $A_u(L)$.

But $A_u(L)/S$ has the canonical power-products of length $\leqq w$ as a (redundant) basis; hence it has a finite order at most $r + r^2 + \cdots + r^w$. Since finally no linear combination of the e_i is of weight $> w$, we see that the different elements of L lie in different residue classes of S, and so

THEOREM 4: *Any nilpotent Lie algebra of finite order can be imbedded in a (nilpotent) linear associative algebra of finite order.*

If one defines w-nilpotence as meaning that all brackets of length $\geqq w$ vanish, one gets (using Poincaré's construction again) the somewhat more precise

[6] By the "length" of any (complex) bracket such as $[[xy]z]$ is meant the number of letters in it, repeated occurrences of the same letter being counted repeatedly.

COROLLARY: *Any w-nilpotent Lie algebra of order r is isomorphic with a Lie algebra of finite matrices of degree at most* $(r^{w+1} - 1)/(r - 1)$, *with coefficients in the same field.*

While if one arranges the basis-elements of $A_u(L)/S$ in order of increasing weight (starting with the adjoined principal unit e_σ of weight zero), one sees—again since multiplying by a monomial increases, and straightening never decreases weight—that

THEOREM 5: *The matrices of the preceding corollary are properly triangular.*

5. **Corollaries of Theorems 4-5.** So far no restriction except commutativity has been placed on the abstract field F underlying the algebras discussed. We shall now specialize to the cases of the real and complex fields, and study the implications of Theorems 4–5 for the theory of finite continuous groups—assuming Lie's Third Fundamental Theorem (but not its converse!).

First let us term a Lie group nucleus "hypercentral" if and only if its infinitesimal algebra is nilpotent.[7]

THEOREM 6: *Every hypercentral Lie group nucleus G is locally isomorphic with a group of matrices whose manifold is homeomorphic in the large with Cartesian r-space.*

PROOF: By definition, the infinitesimal algebra of G is nilpotent—and so by Theorem 5 is isomorphic with an algebra A of properly triangular matrices. It follows by Lie's Third Fundamental Theorem that G is locally isomorphic with the Lie group nucleus generated by these matrices.

But these matrices are properly triangular, and so all logarithmic and exponential series[8] involving them or even polynomials in them (which will also be properly triangular) are actually polynomial—contain only a finite number of non-vanishing terms. Hence by a formal identity of Hausdorff,[9] exponentials of these matrices are a group G^* in the large. Actually, $e^X e^Y = e^{\phi(X,Y)}$ where $\phi(X, Y)$ is a suitable polynomial in X and Y and their alternants. Again, since the logarithmic series is finite, and a fortiori convergent and continuous, the correspondence $X \leftrightarrows e^X - I$ is bicontinuous. Therefore canonical parameters give a one-one representation in the large, and the group G^* has a manifold homeomorphic with the linear r-space of its infinitesimal generators.

COROLLARY 1: *The universal covering group of every hypercentral Lie group is topologically isomorphic in the large with a linear group.*[10]

---

[7] This is equivalent to requiring that its algebraically defined "lower central series" (using the terminology of P. Hall) terminate with the identity.

[8] By e^X we mean $1 + X + (1/2!)X^2 + \cdots$, and by Log $(1 + X)$, the limit of the series $X - \frac{1}{2}X^2 + \frac{1}{3}X^3 - \cdots$, in case it exists. The ideas used are to be found in J. von Neumann, "Gruppen linearen Transformationen," Math. Zeits. 30 (1929), 3–42.

[9] F. Hausdorff, "Die symbolische Exponentialformel in der Gruppentheorie," Leipz. Ber. 58 (1906), Theorem B, p. 29.

[10] By Schreier's well-known theorem that any two simply connected locally isomorphic

COROLLARY 2: *The converse of Lie's Third Fundamental Theorem is valid in the large for nilpotent Lie algebras (with real or complex coefficients).*

Corollary 2 has been proved without restriction by E. Cartan [Fasc. 42 des Sci. Math. (1930), p. 18, bottom].

6. Linear groups with nilpotent algebras.[11] In this section an argument of E. Cartan (communicated by letter) is given, which gives the topology of the most general linear group "of rank zero" (i.e., with nilpotent Lie algebra).

Let L be any nilpotent Lie algebra of finite order r; then L is by §5 the infinitesimal algebra of a simply connected group $\mathfrak{U}$ of strictly triangular matrices—in which the exponential and logarithmic functions are both single-valued. By Schreier's theory of local isomorphisms one concludes that to get the most general Lie group having L for Lie algebra, one should take a general discrete subgroup $\mathfrak{K}$ of the central $\mathfrak{Z}$ of $\mathfrak{U}$, and form $\mathfrak{U}/\mathfrak{K}$.

But since two matrices $X \in \mathfrak{U}$ and $Y \in \mathfrak{U}$ commute if and only if Log Y and Log X commute, $\mathfrak{Z}$ is connected, and even isomorphic with the translation-group of n-space, while $\mathfrak{K}$ is generated by $k \leqq n$ linearly independent matrices $Z_1, \cdots, Z_k$. Taking the Log Z_i, we see that $\mathfrak{K}$ is part of a k-parameter connected normal subgroup $\mathfrak{K}^*$ of $\mathfrak{Z}$, whence $\mathfrak{K}^*/\mathfrak{K}$ is the k-parameter torus group.

Therefore if $\mathfrak{U}/\mathfrak{K}$ is isomorphic with a linear group, $\mathfrak{K}^*/\mathfrak{K}$ corresponds to a group of diagonal matrices whose non-zero terms all have absolute value unity, and which commute with every $X \in (\mathfrak{U}/\mathfrak{K})$. And by a theorem proved elsewhere,[12] the derived subgroup $\mathfrak{U}'$ of $\mathfrak{U}$ generated by Poisson brackets of its infinitesimal generators, can have no elements in common with this.

It follows that by taking suitable generators in $\mathfrak{U}'/\mathfrak{U}$, we can get a subgroup $\mathfrak{G}$ containing $\mathfrak{U}'$, such that $\mathfrak{G} \cap \mathfrak{K}^* = 0$, $\mathfrak{G} \cup \mathfrak{K}^* = L$. But $\mathfrak{K}^*$ being in the central, we see that L must be the *direct* sum of $\mathfrak{G}$ (which is simply connected, by reference to $\mathfrak{U}$) and the toroidal group. This proves

THEOREM 7: *Any group of matrices whose Lie algebra is nilpotent is a direct product of a toroidal group of diagonal matrices and of a simply connected Lie group.*

HARVARD UNIVERSITY.

---

groups are isomorphic in the large. But there exist hypercentral Lie groups simply isomorphic with no linear group even in the purely algebraic sense. Cf. the author's note, Bull. Am. Math. Soc. 42 (1936), 883–888. Also, Cartan has pointed out ("La topologie des groupes de Lie," Hermann, Paris, 1936, p. 18) that the universal covering group of the projective group on one real variable is topologically isomorphic with no linear group —which controverts the conjecture that all simply connected Lie groups are isomorphic in the large with linear groups.

[11] Added March 4, 1937.

[12] "Lie groups isomorphic with no linear group," Bull. Am. Math. Soc. 42 (1936), 882–8. The result used is Theorem 2.